95

JOURNAL OF MATHEMATICS AND PHYSICS

Massachusetts Institute of Technology

Volume IV
January, April, May, July, 1925

EDITORS

FREDERICK S. WOODS, Professor of Mathematics
HARRY M. GOODWIN, Professor of Physics and Electro-Chemistry
FREDERICK G. KEYES, Professor of Physico-Chemical Research

MANAGING EDITOR

CLARENCE L. E. MOORE, Professor of Mathematics

UNIVERSITY

QA1
J87
v.4

CONTENTS

PAGE

Ozone as an Absorbing Material for Radiations in the Atmosphere . 1
By CHARLES FABRY

Note on a Paper of O. Perron 21
By NORBERT WIENER

The Torsion Problem of Curved Beams 33
By PAUL HEYMANS AND W. J. HEYMANS

A Taylor's Expansion of a Determinant 62
By LEPINE HALL RICE

Sommerfeld's Theory of Fine Structure from the Standpoint of General Relativity 65

The Representation of the Haber Equilibrium Data by an Equation 84
By LOUIS J. GILLESPIE

The Electric Currents in a Network 96
By PHILIP FRANKLIN

Determination of the Stresses in a Beam by Means of the Principle of Least Work 103
By WILLIAM HOVGAARD

Determinantal Expression of Multiple Cross and Dot Products of Polyadics in Three Dimensions 130
By LEPINE HALL RICE

Note on the Scattering of X-rays 133
By P. DEBYE

Continued on Page Three

STATE OHIO
UNIVERSITY

CONTENTS—Continued

PAGE

The Weierstrass Approximation Theorem 148
By PHILIP FRANKLIN

The Solution of a Difference Equation by Trigonometrical Integrals 153
By NORBERT WIENER

Note on a Method of Evaluating the Complex Roots of Sixth — and Higher — Order Equations 164
By L. F. WOODRUFF

Minimal Varieties of Two and Three Dimensions Whose Element of Arc is a Perfect Square 167
By C. L. E. MOORE

The Multiple Complement of One or More Polyadics . . . 179
By FRANK L. HITCHCOCK AND LEPINE HALL RICE

Note on De Sitter's Universe 188
By G. LEMAITRE

Note on Quasi-Analytic Functions 193
By NORBERT WIENER

File Multiplication of Ordered Determinants 200
By LEPINE HALL RICE

A Theory of Ordered Determinants with Application to Polyadics . 205
By FRANK L. HITCHCOCK

A New Method in the Theory of Quantics 238
By FRANK L. HITCHCOCK

315225

OZONE AS AN ABSORBING MATERIAL FOR RADIATIONS IN THE ATMOSPHERE

By CHARLES FABRY[1]

The question which I shall treat in the present paper is a part of the very important problem of absorption of radiations in the atmosphere surrounding the earth. This problem is important in astrophysics on account of the fact that all the light which we receive from the stars has traversed our atmosphere; it is important in geophysics because this absorption plays an important part in the atmospheric phenomena; finally, it has its importance in pure physics on account of the remarkable optical properties shown by ozone.

The problem originated in the study of the ultraviolet part of the solar spectrum. It is well known that this spectrum has a great intensity at about the wave length 3100 Angstroms, the intensity decreasing rapidly from that point and practically vanishing a little beyond 2900 A.

A summary study shows that atmospheric absorption plays a large part in this limitation. Indeed, the spectrum of all the stars presents this same limitation, and one would hesitate to say that the radiation of all the stars would be limited in the same way in the shorter wave lengths. Moreover, when the sun sinks towards the horizon, the thickness of the atmosphere which the radiations traverse becomes greater, and the limit of the spectrum recedes towards the longer wave lengths.

After an attentive study of the question, Cornu[2] came to the conclusion more than forty years ago that the absorption was due to a gas existing in the atmosphere in a permanent and very nearly constant manner (thereby excluding water vapor).

A very short time afterwards, Hartley discovered the broad absorption band of ozone in the ultraviolet, and attributed to that gas the limitation of the solar spectrum. However, this

[1]Professor at The University of Paris. Lecturer at the Physics Department of the Massachusetts Institute of Technology, year 1924–1925.

[2]Journal de Physique, lère serie, tome X, page 5, 1881; Comptes Rendus de l'Acad. des Sciences, Vol. CXI, p. 941, 1890.

hypothesis remained very uncertain as only an extraordinarily small amount of ozone was known to exist in the accessible part of our atmosphere. Indeed, finding that the absorption band of ozone roughly coincides with the absorption band of the atmosphere is not a sufficient proof for stating that the two absorbing materials are the same.

The very existence of the radiations of short wave lengths in the radiations *emitted* by the sun has been questioned by Méthe and Lehmann[3] after they had attempted to extend the solar spectrum by experiments which they carried out at great altitudes. Ascending in a balloon to a height of nine thousand meters, they succeeded in obtaining a spectrogram of the sun; in this spectrum they did not detect any change in the extreme limit of the solar spectrum. They concluded that the radiation below 2900 A. *did not exist* in the radiations *emitted* by the sun, and that the limitation was due to an absorption in the solar atmosphere.

I became interested in this question together with Mr. Buisson at Marseilles in 1912. It results from what I said above, that the knowledge at that time of this question was very limited. The absorption by ozone was still a hypothesis; the fact itself of the limitation of the spectrum solely by atmospheric absorption was contested.

We decided that quantitative data on this subject should be secured where only qualitative information was available. This work[4] involved a great number of measurements of intensities of radiations in a continuous spectrum of very low intensity, presenting numerous Fraunhofer lines whose intensity increases very rapidly towards the shorter wave lengths. All those measurements have been made by the method of photographic photometry.[5] This method requires a great deal of care and it has often led to very incorrect results when improperly used. In the present case no other known method could have been employed. Indeed the bolometer or thermopile is not sensitive enough at the required resolving power. Abbot, who is a recognized master in the use of

[3] Sitzungsberichte der Kön. Preus, Akademie der Wiss. 1909, Vol. 1, p. 268.
[4] Fabry & Buisson. Journal de Physique, series 6, t. 2, p. 197, 1921.
[5] The properties of photographic plates are described in: Buisson et Fabry, Les Lois du noircissement des plaques photographiques. Revue d'Optique, t. 3, p. 1, 1924.

those instruments, has not succeeded in carrying out measurements beyond 3400 A. The photoelectric cell has, for a few years, been often used, but it also lacks the necessary resolving power.

Our measurements were planned so as to include:

(1) A quantitative study of the absorption spectrum of ozone, namely the measurement of the coefficient of absorption for the different radiations and from that the determination of the absorption curve.

(2) The measurement of the coefficient of atmospheric absorption for the different radiations and the comparing of the absorption curve in the atmosphere with the absorption curve in ozone; if these two curves should coincide, the atmospheric absorption would certainly be due to ozone and one would thereby be able to calculate the thickness of ozone existing in the atmosphere.

(3) The energy distribution curve in the spectrum of solar radiation as *received* through our atmosphere when the sun is near the zenith.

(4) Having obtained this curve and knowing the atmospheric absorption, the energy distribution curve of the radiation *emitted* by the sun: will this curve present a great drop when approaching the limits which can be observed or on the contrary will the radiation be maintained up to that limit?

Study of the Absorption by Ozone

The measurements which had been made by Meyer[6] and by Krüger and Moeller[7] by the photoelectric method did not cover the entire region which we wanted to study. Moreover, their results disagreed; indeed in the region $\lambda = 3000$, their values for the absorption constant are in the ratio of 1 to 4.

With Mr. Buisson, I redetermined (1912) those constants by the photographic method. The radiation of a quartz mercury lamp is made to travel through a tube which is closed by two quartz plates. The light is analyzed in a spectrograph with two quartz prisms, and the intensity curve can be modified in a known ratio by means of diaphragms of varying area which are placed after the second prism. On the same photographic plate with the

[6] Annalen der Physik. Series 4, t. 12, p. 849, 1903.
[7] Physikalische Zeitschift., t. 13, p. 729, 1912.

same length of exposure, an exposure is first made with full aperture, the tube being filled with a known mixture of oxygen and ozone, and afterwards a series of exposures, the tube containing pure oxygen. The *optical densities* of the different images allow one to calculate the coefficient of transmission for the different radiations and therefrom to obtain this factor for a layer of pure gaseous ozone of unit thickness. All the measurements of the *densities* of the photographic plates have been made by means of our microphotometer.[8]

Various coefficients are used to express the absorption of a substance for a given radiation. The coefficient which we use is defined in the following manner:

Let x be the thickness expressed in centimeters of pure gaseous ozone; I_0 the intensity of the incident radiation; I the intensity of transmitted radiation. The value of I as a function of x is given by the following equation:

$$I = I_0 10^{-ax}$$

where a is the coefficient of absorption.

When x equals one centimeter, this expression becomes:

$$a = \log_{10} \frac{I_0}{I}.$$

Fig. 1 gives the values of a as a function of the wave length.

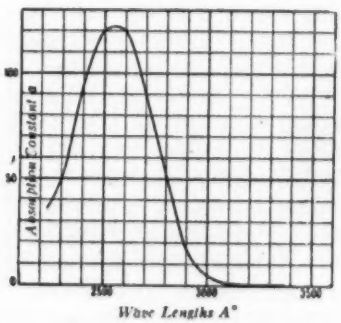


Fig. 1

[8]Journal de Physique, Series 5, t. 9, p. 31, 1919 and Revue d'Optique, t. 1, p. 1.

One should note the large value of the maximum of a which reaches 123 (for about $\lambda = 2550$). In this region of the spectrum the ozone has a really extraordinary absorbing power; a layer of pure gaseous ozone 25 μ in thickness would reduce by one-half the intensity of the radiation, and if the ozone could be reduced in volume so as to have the density of silver, the same result would be produced by a layer of 5 $m\mu$. The absorption of ozone in this region is really *metallic* absorption.

This wide absorption band of ozone coincides with the region of the spectrum which decomposes the ozone O_3 into oxygen O_2. The region which interests us for the study of the solar spectrum is the part which lies at the side of the long wave lengths with respect to the maximum. The following empirical relation expresses well the experimental values between 2900 and 3200 A.

$$\log a = 17.58 - 0.00564 \lambda.$$

Spectrograph for the End of the Solar Spectrum[9]

When the ultraviolet end of the solar spectrum is photographed with an ordinary spectroscope, one finds this ultraviolet end obscured by a fog which limits the region where observations can be made. This fogging is produced by the stray light belonging to the other regions of the spectrum whose intensity is extremely great when compared with the intensity of the extreme region. Stray light is originated by diffusion in all solid media and by reflection at all the surfaces separating two media. In order to *purify* the spectrum a second dispersing system must be used, at right angles to the first one, rejecting sideways the stray light outside of the spectrum.

Fig. 2 gives a perspective of the apparatus. It consists of two spectrographs placed in series in two planes at right angles. Both of the spectrographs employ two quartz prisms of 60°, one right and one left handed, and two quartz lenses. In the first spectrograph FLPP'L', both of the quartz lenses have a focal length of 1 meter; in the second spectrograph SLPP'L' those lenses have a focal length of respectively 25 and 17 centimeters. A spectrum perfectly free from stray light is thereby obtained in B and

[9] Ch. Fabry et H. Buisson. Study of the Ultra-violet End of the Solar Spectrum; Journal de Physique, series 6, t. 2, p. 197, 1921.

it becomes possible to make exposures of one hour of the solar spectrum without any trace of fogging.

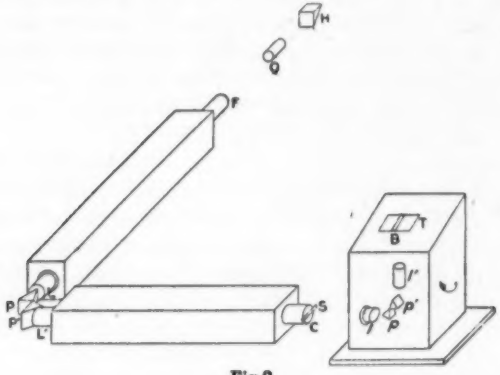


Fig 2

Another difficulty is the fact that in the region of the spectrum which we want to study, the variation of intensity is so considerable that one single exposure only gives results which may be used in an extremely small interval. An exposure of one minute is correct for the region 2936, but such an exposure is of no use after 2975, and for 3100 the right exposure is only of 1/1000 of one second. The following method has overcome the difficulty arising from the short latitude of the photographic plate.

On a part of the slit S on which the first spectrum is projected (Fig. 2) a system of two filters very nearly neutral is placed as shown in Fig. 3. The radiations of a wave length below 2990

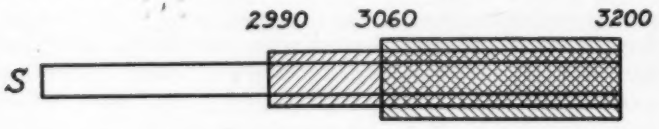


Fig. 3

do not undergo any absorption; the region 2990−3040 is reduced to 1/100 of its intensity by the interposition of a filter, and the

region 3040−3200 undergoes the absorption due to the two filters which transmit only about 1/2000 of the intensity.

Those filters consist of the gelatine film taken from the glass of a photograph plate which has been developed, fixed and intensified[10] by means of mercuric chloride.

Measurement of the Coefficient of Absorption of the Atmosphere for the Various Radiations

The ordinary method for measuring the coefficient of absorption consists in comparing the intensities at a particular wave length, both with and without interposition of the absorbing medium. This cannot be done for the atmosphere as its removal is impossible; however it is possible to vary the thickness by analyzing the solar light as the sun moves from the zenith. This method might be called the Bouguer-Langley method.

Let I_0 be the intensity of a solar radiation before it enters the atmosphere. If the sun is at the zenith, this radiation reaches the earth with a reduced intensity I_1. The coefficient of absorption m for this radiation is defined by the following equation:

$$m = \log_{10} \frac{I_0}{I_1} .$$

If the sun is at a zenithal distance ζ, the path[11] of the radiation in each layer has to be multipled by sec. ζ and the intensity I which reaches the earth is given by:

$$\log I = \log I_0 - m \text{ sec. } \zeta.$$

The intensity I is measured at different times of the day, and the curve is traced, taking as abscissæ the values of sec. ζ and as ordinates the values of log I. Those measurements of intensities (or rather the *ratios* of the intensities) of a particular wave length at different hours of the day are made photographically, the exposures of the same length of time being made on the same plate. Moreover, at mid-day, a graded series of exposures is made by means of diaphragms or series of filters to obtain the characteristic of the plate.

[10] This intensification is in order to suppress the selective absorption of ordinary photographic plates which when simply developed and fixed, consist of colloidal silver. See Revue d'Optique, vol. 3, p. 1, 1924.

[11] Neglecting the curvature of the earth and of the atmospheric layers, this being permissible when the sun is not too near the horizon.

The absorption in the atmosphere must be attributed to three distinct causes:

1. The main absorption which increases rapidly as the wave length decreases and is due to the gas which we tried to identify as ozone.

2. The diffusion due to the molecules of the air (Lord Rayleigh's phenomena) which produces a weakening which can be calculated for each radiation.

3. The absorption produced by the moisture and other substances in the atmosphere; this is practically independent of λ in the small interval which we consider, but it might vary from time to time in some way unknown.

Let us call m the coefficient of absorption of the atmosphere due to the unknown substance, and β the coefficient of absorption due to molecular diffusion when the sun is at the zenith. These two quantities depend only on λ. I_0 being the intensity outside of the atmosphere, the intensity I at the surface of the earth, when the sun is at a zenithal distance ζ, is given by the equation:

$$\log I = \log I_0 - (m+\beta) \sec \zeta - \delta$$

in which δ is a number representing the coefficient of absorption due to moisture at the time of the experiment, nearly independent of λ but varying irregularly during the day.

For another radiation λ' we have at the same instant:

$$\log I' = \log I'_0 - (m'+\beta') \sec. \zeta - \delta$$

Eliminating δ we obtain:

$$\log I - \log I' = K - (m-m'+\beta-\beta') \sec \zeta$$

where K is a quantity independent of the time.

We take for λ', the radiation of longest wave length in the region studied ($\lambda' = 3143$) for which the absorption is slight. For each of the other radiations we can draw a diagram in which the values of sec ζ are taken as abscissæ and the values of $\log I - \log I'$ as ordinates. The curve which is obtained must be a straight line, the slope of which is equal to $(m-m'+\beta-\beta')$.

Table 1 gives the values of $\log I - \log I'$ (with the addition of an arbitrary constant in order to avoid negative values) at the different hours of June 7, 1920.

TABLE 1

Hour	Sec. ζ	Wave Lengths in Angstroms										
		2922	2931	2936	2946	2956	2963	2997	3022	3052	3104	3143
8.22*	1.526					0.85	1.26	2.94	3.50	4.42	4.97	5.22
8.39*	1.440				0.51	1.09	1.46	2.07	3.64	4.46	5.00	5.22
8.56*	1.367				0.68	1.26	1.66	3.14	3.70	4.47	5.02	5.22
9.16*	1.295			0.43	0.91	1.49	1.79	3.29	3.76	4.55	4.98	5.22
9.38*	1.232		0.31	0.59	1.09	1.63	1.94	3.37	3.79	4.54	4.95	5.22
10.02*	1.177		0.49	0.78	1.26	1.79	2.14	3.44	3.85	4.55	4.98	5.22
10.31*	1.128	0.22	0.62	0.93	1.38	1.90	2.23	3.48	3.92	4.58	5.00	5.22
11.25†	1.077	0.37	0.78	1.11	1.53	2.08	2.33	3.58	4.00	4.62	5.02	5.22
1.22†	1.118	0.26	0.69	0.99	1.45	1.98	2.31	3.53	3.95	4.61	5.04	5.22
1.52†	1.165		0.53	0.81	1.30	1.83	2.16	3.48	3.90	4.59	5.01	5.22
2.24†	1.237		0.34	0.59	1.10	1.65	2.02	3.39	3.84	4.59	5.03	5.22
2.44†	1.295			0.48	0.95	1.52	1.95	3.31	3.78	4.57	5.03	5.22
3.04†	1.367				0.75	1.36	1.67	3.23	3.72	4.52	5.04	5.22
3.20†	1.435				0.55	1.15	1.55	3.13	3.69	4.48	5.01	5.22
3.46†	1.573					0.88	1.22	2.99	3.58	4.45	5.03	5.22

*a. m. † p. m.

In Fig. 4 those observations have been put in the form of a graph. It is seen that for each radiation the points follow very closely to a straight line and that the slope of these straight lines increases rapidly as one goes towards the shorter wave lengths.

The slope of each of these lines gives the value of $(m-m'+\beta-\beta')$. On the other hand the values of β and β' (molecular diffusion) are known; the values of $(m-m')$ can therefore be calculated.

These results have now to be compared with the absorption by ozone. If we assume that it is gaseous ozone which produces the atmospheric absorption and represent by x the total thickness of it (reduced to atmospheric conditions) through which a ray coming from the zenith must pass, we have $\beta-\beta'=(\alpha-\alpha')\,x$ where α and α' are the coefficients of absorption by ozone respectively for the radiation used and for the radiation 3143. These coefficients are known. The values of $\beta-\beta'$ given by solar observation must therefore be proportional to the values of $\alpha-\alpha'$ obtained with ozone, the proportionality constant being the total thickness of ozone expressed in centimeters contained in the atmosphere.

In Table 2 these calculations for one day of observation have been recapitulated. The values given in column P are the *slopes*

TABLE 2

λ	β	$\beta-\beta'$	α	$\alpha-\alpha'$	P	$m-m'$	x
2931	0.47	0.11	11.2	10.5	3.15	3.04	0.290
2936	0.46	0.10	10.5	9.8	3.16	3.06	0.312
2946	0.46	0.10	9.3	8.6	2.78	2.68	0.312
2956	0.45	0.09	8.1	7.4	2.66	2.57	0.347
2963	0.45	0.09	7.4	6.7	2.37	2.28	0.340
2997	0.43	0.07	4.7	4.0	1.32	1.25	0.309
3022	0.42	0.06	3.4	2.7	0.96	0.90	0.333
3052	0.40	0.04	2.3	1.6	0.44	0.40	0.250
3143	0.36	0.00	0.7	0	0	0	..

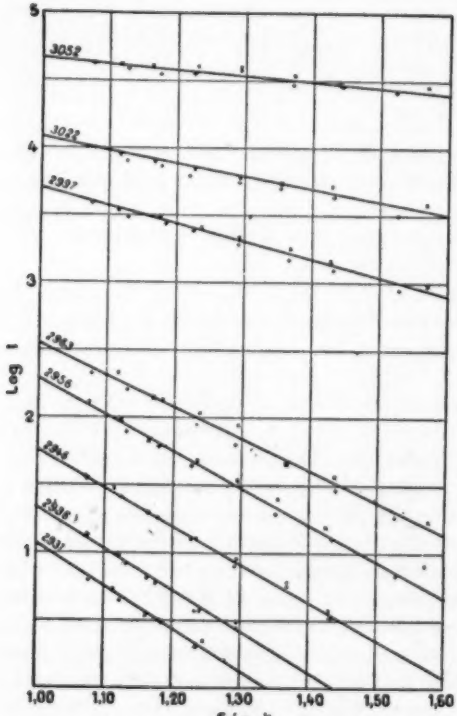


Fig. 4

of the straight lines of the figure. The last column gives the values of x calculated from $x=\frac{\alpha-\alpha'}{\beta-\beta'}$.

This quantity is practically constant within the degree of accuracy which can be expected in measurements of this sort. The first and last radiations give uncertain results: the first because of the difficulty in measuring, due to its very feeble intensity; the second because its absorption is not very great and therefore x is the ratio of two very small numbers.

We may conclude that *it is actually ozone which produces the absorption in the ultraviolet in the atmosphere*. Moreover, the total thickness of this gas is about 3 mm. Table 3 gives the results for three other days of observation.

TABLE 3

λ	*May* 13	*May* 20	June 7
2931	—	0.34	0.29
2936	0.30	0.34	0.31
2946	0.33	0.32	0.31
2956	0.33	0.33	0.35
2963	0.35	0.35	0.34
2997	0.33	0.36	0.31
3022	0.38	0.34	0.33
3052	0.30	0.33	0.25
Adopted Value	0.33	0.34	0.32

The Solar Spectrum without Atmospheric Absorption

The relative intensities at a given number of points in the solar spectrum were first determined under the conditions in which these radiations reach us through our atmosphere. Those intensities have to be expressed correctly in terms of energy: if I represents the intensity in the vicinity of the radiation λ, $\lambda d\lambda$ is proportional to the energy received during a unit of time by a unit of surface in the form of radiations comprised in the interval between λ and $\lambda+d\lambda$. In order to eliminate all influences of selective sensitivity by the photographic plates and of unequal dispersion of the prisms in the different parts of the spectrum, we proceeded by way of comparison with the radiation from the positive crater of an electric arc which can be considered as a black body at 3750°C.

As the solar spectrum presents in this region numerous strong Fraunhofer lines, we have chosen a certain number of points where those rays are less numerous and less strong. The measurements, therefore, have been made as much as possible in the parts of the continuous background of the spectrum although the regions between the strong Fraunhofer lines contain a large number of fine absorption lines.

Knowing the decrease in intensity due to the atmosphere, we can then calculate the relative intensities of the radiation *emitted* by the sun. In Table 4 the results obtained during the observations made on June 7, 1920, are recapitulated:

TABLE 4

λ	Intensity at Sea Level	Coefficient of Absorption	Intensity emitted by the Sun
3143	22400	0.84	15.5
3104	15100	0.99	14.7
3092	10200	1.40	29.6
3022	2700	1.77	19.8
2997	1320	2.22	21.8
2963	132	3.10	16.6
2956	76	3.33	16.2
2946	25	3.73	13.5
2936	11	4.12	14.1
2931	5.5	4.36	12.6
2922	2.2	4.82	14.9
2917	0.87	5.08	10.9
2912	0.30	5.39	7.4
2906	0.04	5.78	2.5
2898	0.02	6.36	4.5

In the first column the wave length is given; in the second, the intensity of the radiation as *received* at the level of the sea when the sun is at the zenith; in the third, the coefficient of total absorption of the atmosphere (logarithm of the ratio of the incident to the transmitted intensity); in the fourth, the intensity *emitted* by the sun. All these intensities are given in arbitrary units and only the ratios of the numbers of the same column are of interest.

Regarding this last table, following should be noted:

1. The enormous decrease of intensity of the radiation received at the surface of the earth as one approaches the limit of the spectrum. The intensity at 3143 is approximately 10 times the intensity at 2898.

2. The enormous value of the atmospheric absorption as one approaches the limit. At 2898 the atmosphere allows only one two-millionth of the incident intensity to pass through.

3. The fluctuations in the spectrum of the intensities of the emitted radiation are readily explained by the Fraunhofer lines: there is a small decrease towards the short wave lengths as should be expected for a radiation which approximates to black body conditions. The sudden decrease near the limit which would explain the limitation of the spectrum independently of atmospheric absorption does not exist. The atmospheric absorption suffices perfectly to explain the phenomena which have been observed and there is no indication of absorption in the solar atmosphere (except the one which produces the Fraunhofer lines).

Variations in the Quantity of Ozone

The thickness of ozone contained in the atmosphere can be measured daily by the method just described. Small variations are found from day to day. On a total number of fourteen days of observation, Mr. Buisson and I have found variations from 2.85 to 3.35 mm. They appear irregular, sometimes rather rapid, without relation to meteorological conditions.

Location and Origin of the Atmospheric Ozone

The preceding observations fail to give information concerning the location of said ozone. If the layer of 3 mm. was uniformly distributed in the atmosphere, the proportion of ozone in volume would be 4×10^{-7}, or in weight, 60 mgrs. per 100 kgms. Chemical analysis of the atmosphere gives a much smaller proportion; Mr. Lespieau[12] found 3 mgrs. of ozone per 100 kgms. of air at sea level, and the same proportion on the top of Mont Blanc (4800 m. altitude). R. J. Strutt[13] has shown by means of a more reliable experiment than chemical analysis that the proportion of ozone in the lower atmosphere is very small. He observed the ultra-violet radiation emitted from an artificial source through a thick layer of air and he found that the mercury rays 2536 were still transmitted in a very appreciable manner over a distance of 6.5 km. If the ozone were uniformly distributed, this thickness

[12]Comptes-Rendus de l'Academie de Sciences.
[13]Proc. Royal Soc. Ser. A. t. XCIV, p. 260, 1918.

of ozone would transmit only a fraction 1×10^{-33} of this radiation, a thickness of 1 km., reducing the intensity to 10^{-6}.

We must therefore conclude that the greatest part of the ozone lies in the very high atmosphere outside of the regions which are accessible to us. The presence of ozone in the very high regions, perhaps over 50 km., can be readily explained. The oxygen is transformed into ozone by the radiations of very short wave lengths (below 1900). It is not unlikely that these short wave lengths exist in the radiations emitted by the sun, since, in the spectral region which we can study, no rapid fall in intensity is observed. These radiations will produce an ozonization of the atmosphere, but being strongly absorbed by the oxygen, they cannot penetrate deeper and are thereby only active in the uppermost layers. The radiation 1850 is completely absorbed by 10 m. of air at ordinary pressure, and does not penetrate beyond the layer at 40 km. of altitude. The radiations in the Schumann region are still more strongly absorbed. On the other hand, the ultraviolet radiations of the region 2000−3000 A. decompose ozone. Between these two inverse actions, a condition of equilibrium establishes itself, and the quantity of ozone which exists must depend on the relative intensities in the two regions of the spectrum. It would be very interesting to follow these variations during an entire solar period and thereby to study the rapid variations detected by Buisson and myself. A simultaneous study in different places on the earth would be useful in order to eliminate the local disturbances.

The Solar Spectrum in the Region 2200−2000

The absorption by ozone has its maximum at about 2600; it becomes extremely small for the wave lengths below 2200. If ozone were the only absorbing material in the atmosphere, we would receive from the sun an appreciable proportion of the radiations which are emitted. However, those radiations are absent in our observations.

This last question has been examined by Mr. Duclaux and Mr. Jeantet,[14] and they have found a very strong absorption due to ammonia in this region. As this gas is always present in the

[14] Journal de Physique. t. 4, p. 115, 1923.

lower atmosphere, Mr. Duclaux thinks that its absorption suffices to explain the absence of the radiations which have not yet been absorbed by ozone. As ammonia is produced at the surface of the earth, it can be absent from the higher atmosphere, and therefore the radiations which have escaped the absorption by ozone might reach the summit of the high mountains. This question has been examined by three of my students, Messrs. Dejardin, Chalange, and Lambert, who have tried to detect radiations in the region 2200−2000 in the solar spectrum at the summit of Mont Blanc.[15] For such observations all stray light in the spectrum had to be entirely avoided. For that reason a double spectrograph of a different type from the one described above was used. They detected no trace of said radiations. Therefore the absorption in the atmosphere must be produced by another gas than NH_3.

Oxygen may possibly suffice to explain this absorption; its very strong absorption band in the Schumann region might very well extend for very great thicknesses up to the region 2200 and beyond. Further experiments would be necessary to settle this question.

Absorption Bands of Ozone in the Region 3300

At the other end of the wide absorption band of ozone, namely, towards the longest wave lengths, a series of narrower absorption bands are found to exist from 3150 to 3400 approximately. They constitute a group perfectly characteristic of ozone.

Fowler and Strutt[16] have shown that these bands exist in the spectrum of the stars of type *A*, such as Sirius, which present only a few Fraunhofer lines. The observation is more difficult on the sun whose spectrum is covered with absorption lines in that region; the bands of ozone may, however, be observed when the sun is very near the horizon. The values of the coefficients of absorption in this region have been recently measured by one of my students, Mr. Lambrey, who has employed the photographic method described above. Fig. 5 gives the absorption curve. This curve allows the calculation of the thickness of ozone which is necessary to render the bands visible; one finds that for a star

[15]Comptes-Rendus de l'Acad. des Sciences. t. 177, p. 757, 1923, also: Revue d'Optique. t. 3, p. 277, 1924.

[16]Proc. Royal Soc. Series A. t. XCIII, p. 577, 1917.

at the zenith, the absorption by 3 mm. of ozone is too small to make the bands easily observable. They should become easy to observe for a thickness four to five times greater, and this is well in accordance with our observations. A precise numerical comparison is not possible as we do not yet possess the values of the coefficient of absorption of the atmosphere.

Mr. Dufay[17] has shown that the absorption spectrum of ozone can be very easily observed in the spectrum of the sky light near the zenith at the end of twilight. The light then comes from the diffusion by the high layers of the atmosphere which are illuminated by solar light which is crossing very obliquely the ozonized air without having to pass obliquely through the low layers.

Absorption Bands of Ozone in the Yellow and Red Region of the Spectrum

Ozone also possesses an absorption spectrum in the visible region discovered in 1882 by Chappuis.[18] The absorption is small and can only be detected for very great thicknesses of gas with a large ozone content. The gas is then sufficiently absorbing so as to become of a blue color. The same phenomenon occurs for liquid or solid ozone.

The values of the absorption coefficients have been determined in my laboratory by Mr. Colange by the photographic method. Fig. 6 gives the absorption curve. The values of the absorption coefficients are very small; the maximum being about .05 for $\lambda = 6000$ (compare with the value 123 for $\lambda = 2550$). A layer of 6 cm. of pure gaseous ozone would be required to reduce the intensity by one half.

It is therefore not surprising that these bands of the visible spectrum are not observed in the solar spectrum. Nevertheless, the absorption by the 3 mm. of ozone is not entirely negligible; for the sun at the zenith and for $\lambda = 6000$ it reaches about 4 per cent.

The values of the coefficient of absorption in the visible part of the spectrum have been measured with great accuracy on

[17] Comptes-Rendus de l'Acad. des Sciences, t. 176, p. 1290, 1923.
[18] Annales de l'Ecole normale superieure, t. 11, p. 137, 1882. also: Journal de Physique. Serie 2, t. 1, p. 494, 1882.

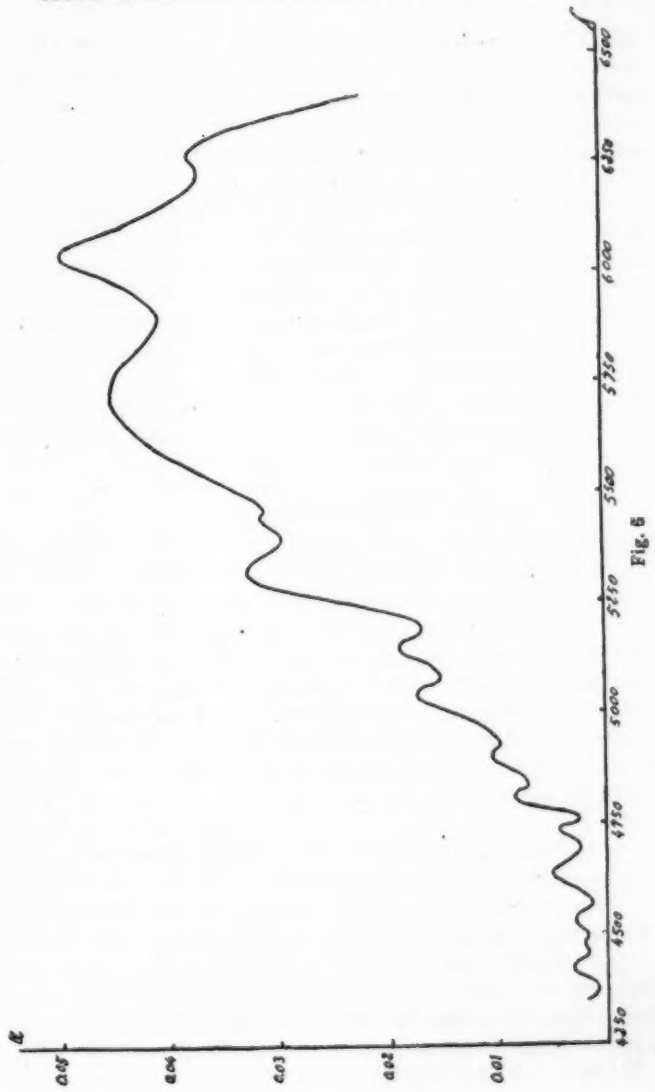

K
0.05
0.04
0.03
0.02
0.01
4250
4500
4750
5000
5250
5500
5750
6000
6250
6500

Fig. 5

Mt. Wilson by Fowle[19] who found that the molecular diffusion, the theory of which is due to Lord Rayleigh, nearly succeeds in explaining the absorption during a perfectly clear day. The question has been recently re-examined by Mr. Jean Cabannes,[20] one of my former students, who is now professor at the University of Montpellier.

Mr. Cabannes has, for several years, devoted a great deal of study to the question of molecular scattering. The Rayleigh theory is only a first approximation which does not take into account the anisotropy of the molecules. Using his own results, Cabannes recalculated the absorption by diffusion and found his results over most of the spectrum in good accord with the values of Fowle; however, in the region 6000 the observed absorption is decidedly greater than the calculated values. This is exactly the absorption region of ozone, and the accord again becomes very close when this absorption is taken into account.

Infra-red Bands

The ozone has also a series of absorption bands in the infra-red region between 5μ and 10μ. An important part in the atmospheric phenomena has been sometimes attributed to these absorption bands where they would act as a " trap " for the radiations and stop a part of the earth's radiations. Unfortunately, we have absolutely no information on the width of these bands, nor the values of the coefficient of absorption and it is therefore impossible to decide whether the supposed effect is real. Mr. Chalange has undertaken in my laboratory a series of measurements on this region of the spectrum.

Possibility for Study in the Very High Atmosphere

It would be interesting to secure exact data regarding the height at which the ozone lies. We arrive to the conclusion that this ozone lies in the high layers only because we do not find it in the lower layers, but it would be of great interest to locate it more definitely even in an approximate manner. There is little hope of obtaining this information by studying the layers which

[19] Astrophysical Journal. t. 38, p. 392, 1913, and t. 40, p. 635, 1914.
[20] Comptes-Rendus de l'Acad. des Sciences, t. 179, p. 191, 1924.

we can reach ourselves. Sounding balloons might perhaps bring some results or gun shells and rockets.

Information might be obtained in the more simple way by observation of the law of absorption as the sun approaches the horizon. If ζ is the apparent zenithal distance of the sun, we have assumed that the thickness of the path through the atmosphere increases as sec. ζ. This would be perfectly correct if the earth and the atmosphere were plane and the refraction neglected. Those assumptions are legitimate for the values of ζ which are not very great, namely, smaller than 45°. But if ζ approaches 90° the law of increase of the thickness and consequently the law of decrease of the intensity depends upon the altitude of the layer causing the absorption. One can readily see by Fig. 7, where the curvature of the layers have been represented, that the incidence becomes less and less oblique as the altitude of the layers increases.

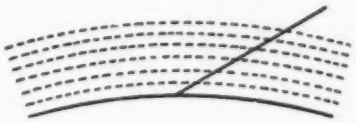


Fig. 7

We can calculate the law of absorption as a function of the zenithal distance for the layers of different altitudes. Then if we had the directly observed law for those layers we would be able to compare them with the different theoretical laws and an indication concerning the altitude of the layer of ozone would thereby be obtained.

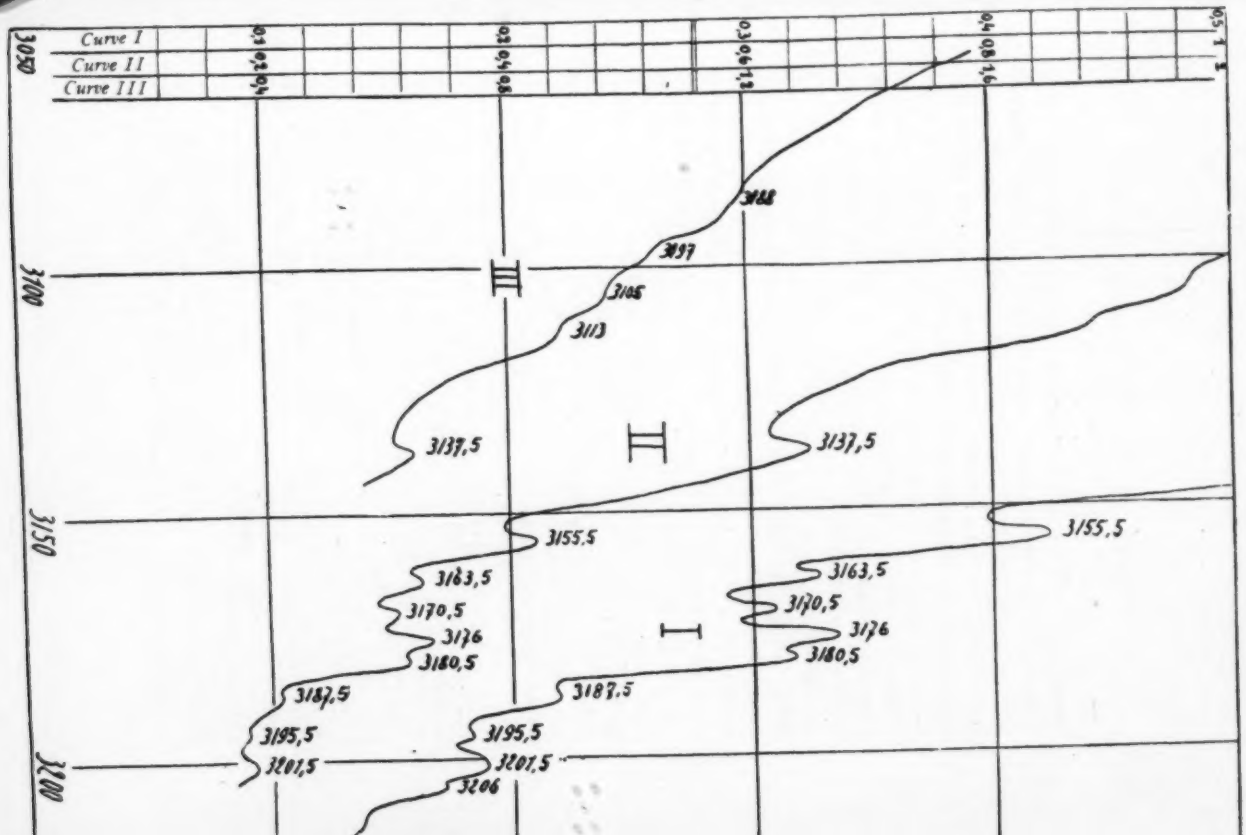
To do this, the solar spectrum must be observed until the sun reaches the horizon. This can only be done effectively on a high mountain, as the great absorption, due to the low layers which are filled with dust and traversed under grazing incidence at sunset, must be avoided.

Such observations were tried in August, 1924, at the top of Mont Blanc by Messrs. Dejardin, Chalange and Lambert. Their photographic plates are at present being measured.

Finally, it has been suggested from different sides (especially by R. W. Wood) that the quantity of ozone might decrease, due to spontaneous decomposition, in the polar regions at the end of

the long night. With the ozonizing radiation cut off, the ozone might entirely disappear. The spectrum of the stars could then be extended beyond the ordinary limit. However, the result of these difficult observations are a priori very uncertain. Indeed, we know nothing about the spontaneous decomposition of ozone in a free space at low temperature without catalytic action of the walls of the container. It is also possible that the atmospheric circulation would bring back to the polar region layers which have been ozonized in other parts of the planet. It would, however, be of the greatest interest if we could reach the strong magnesium line 2802 in the spectrum of the stars.

M. I. T., Physics Department
November, 1924.


Curve I
Curve II
Curve III
3050
3100
3150
3200
0,3 0,4 0,4
0,3 0,4 0,6
0,3 0,6 1,2
0,4 0,8 1,6
0,5 1 3
III
II
I
3188
3197
3205
3213
3137,5
3137,5
3155,5
3155,5
3163,5
3163,5
3170,5
3170,5
3176
3176
3180,5
3180,5
3187,5
3187,5
3195,5
3195,5
3201,5
3201,5
3206

Curve giving the coefficient of absorption α as a function of the wave l

In abscissae λ : *1 division represents 10 Angstroms*

In ordinates α
- *Curve I* 5 divisions represent $\alpha = 0.1$
- *Curve II* 5 divisions represent $\alpha = 0.2$
- *Curve III* 5 divisions represent $\alpha = 0.4$

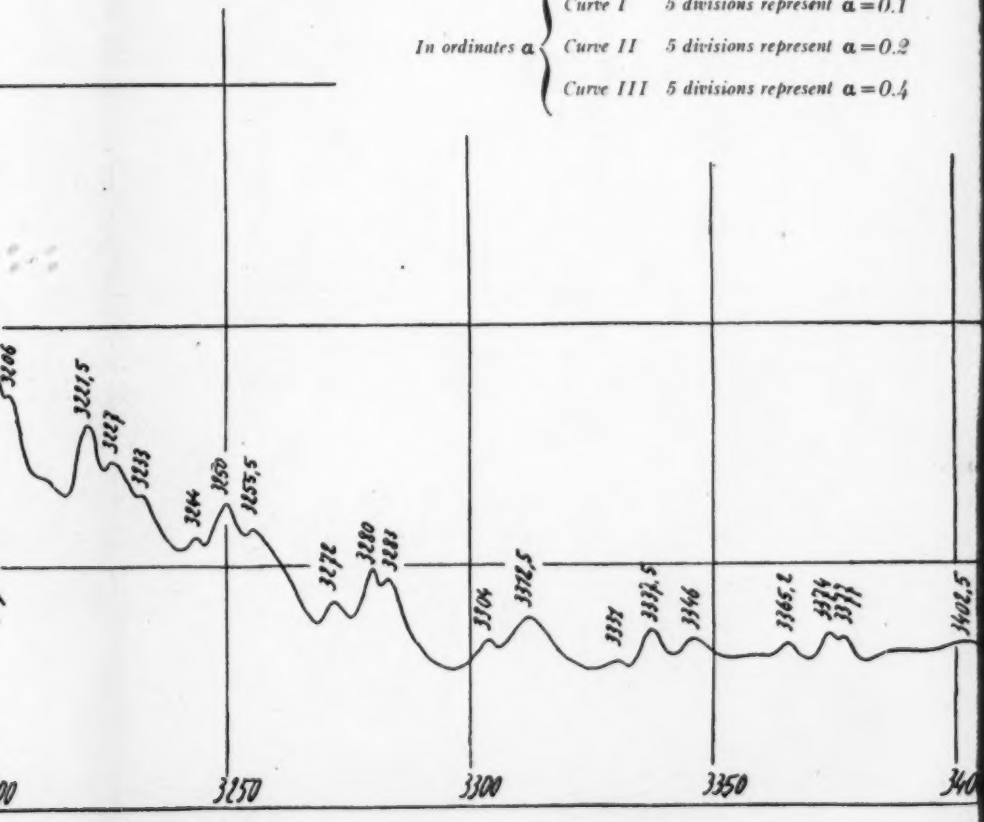


Fig. 5

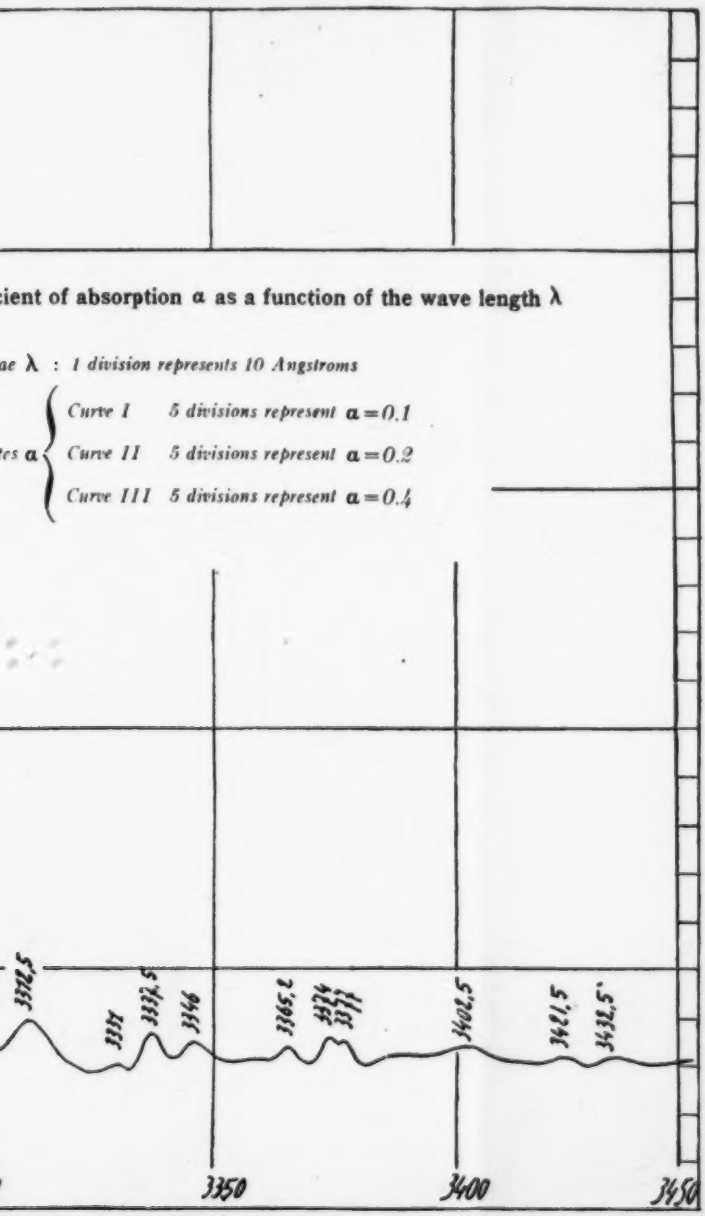
Curve giving the coefficient o

In abscissae λ :

In ordinates α

3155,5
3163,5
3170,5
3176
3180,5
3187,5
3195,5
3201,5
3206
3221,5
3227
3233
3244
3250
3255,5
3272
3280
3283
3304
3312,5
3150
3200
3250
3300

Fig. 5


cient of absorption α as a function of the wave length λ
ae λ : 1 division represents 10 Angstroms
tes α
Curve I 5 divisions represent α = 0.1
Curve II 5 divisions represent α = 0.2
Curve III 5 divisions represent α = 0.4
3372,5
3337
3353,5
3346
3365,2
3374
3393
3402,5
3421,5
3432,5
3350
3400
3450

NOTE ON A PAPER OF O. PERRON

By NORBERT WIENER

§1. **The Generalized Dirichlet Problem for Continuous Boundary Values.** In the last few years, the theory of the first boundary value problem of potential has been brought to a very high degree of completeness and definitiveness by the independent work of a considerable number of investigators in Europe and America.[1] In this development there have been three principal stages or moments: (1) the envisagement of the first boundary value problem (which, as Lebesgue has shown, need not always admit a solution) as a special case of a more general linear problem which is always soluble for continuous boundary conditions; (2) the precise determination of the types of boundaries for which the solution of this more general problem for any continuous boundary condition is *ipso facto* a solution of the first boundary-value problem in the classical sense; (3) the extension of the solution to very general cases of discontinuous boundary conditions. Stage (1) has been clearly outlined by Lebesgue, and has been developed in detail by Bouligand and Wiener, who have likewise obtained what may be regarded as a substantially complete theory of stage (2). In stage (3) the furthest developments thus far obtained are those of Evans and of Wiener, who make use of that extremely powerful tool, the Daniell integral.[2]

[1] Cf. H. Lebesgue, *Sur le problème de Dirichlet*, Rend. di Palermo, v. 24, 1907; note in C. R. des séances de la Soc. Math. de France, 1913, p. 17; note in C. R., Jan. 21, 1924 (pp. 349–354); G. Bouligand, *Sur le problème de Dirichlet harmonique*, C. R., Jan. 2, 1924, pp. 55–57; *Domaines infinis et cas d'exception du problème de Dirichlet*, C. R., March 24, 1924, pp. 1054–1057; *Sur les principes de la théorie du potentiel*, Bull. Sci. Math., v. 48, July, 1924, pp. 227–232; O. D. Kellogg, *An Example in Potential Theory*, Proc. Am. Acad. Arts and Sciences, v. 58, No. 14, June, 1923; G. E. Raynor, *Dirichlet's Problem*, Annals of Math., March, 1922; S. Zaremba, Bull. de l'Ac., des Sc. de Cracovie, July, 1909; G. C. Evans, *Problems of Potential Theory*, Proc. Nat. Acad. Sci., v. 7 (1921) pp. 89–98; H. B. Phillips and N. Wiener, *Nets and the Dirichlet Problem*, this Journal, v. 2 (1923) pp. 105–124; N. Wiener; *Certain Notions in Potential Theory*, this Journal, v. 3, No. 1, January, 1924; *The Dirichlet Problem*, this Journal, v. 4, No. 3, April, 1924; *Discontinuous Boundary Conditions and the Dirichlet Problem*, Trans. Am. Math. Soc., v. 25 (1923), pp. 307–314; O. Perron, *Eine neue Behandlung der ersten Randwertaufgabe für* $\Delta u = 0$, Math. Ztschr., v. 18, Heft 1/2.

[2] P. J. Daniell, *A General Form of Integral*, Annals of Math., (2) 19, pp. 279–294 (1917–1918).

While the paper of Perron has reached the limit of generality and completeness in none of the three aspects of the Dirichlet problem, and while it is substantially contemporaneous with a much shorter paper of Kellogg in which most of its results are contained, it is distinguished by a peculiar elegance and simplicity. Many of its methods are capable of a development far beyond the confines of the actual paper. It is the purpose of the present note to carry out this development, and to show how it is possible to erect on Perron's foundation a theory substantially equivalent to that of Bouligand and Wiener.

In the course of this paper, however, we shall consider a case of the Dirichlet problem somewhat different from that of Perron. The change is in no way essential, but is introduced to eliminate the more or less accidental complications arising, on the one hand, from the distinction between the different sheets of a Riemann surface of space and, on the other, from the behavior of a logarithmic potential at infinity. Perron discusses a plane-bounded set of points G consisting entirely of interior points. His G may lie on a many-sheeted Riemann surface, but contains no branch point in its interior. We shall consider, on the other hand, a three-dimensional bounded set of points G, consisting entirely of interior points, and nowhere overlapping itself. Again, Perron states his boundary-value problem as follows: *On the boundary R let a bounded function f be given. Let $\underline{f}$ and $\overline{f}$ be its inferior and superior functions.*[3] *To find a function u on the closed point set $G+R$ which fulfils the following conditions:*

I. *On the boundary R, the inferior and superior functions satisfy the inequalities*

$$\underline{f} \leq \underline{u} \leq \overline{u} \leq \overline{f};$$

II. *u is harmonic in G.*

Now this form of putting the Dirichlet problem says at once too much and too little: too much, because it goes beyond the case of continuous functions f; and too little, because for all but the simplest discontinuous boundary conditions, I is not strong

[3] *I.e.*, $\underline{f}(P)$ is the smallest limit which can be approached by $f(P_n)$ as P_n approaches P. Likewise, $\overline{f}(P)$ is the largest limit which can be approached by $f(P_n)$ as P_n approaches P.

enough to secure the uniqueness of u. We shall therefore confine our attention for the present to continuous boundary conditions f, and shall reserve for the end of the paper the more general case.

If K is a sphere which with its interior lies entirely within G, and if v is a function defined over $G+R$, then Perron understands by $M_K v$ a function which coincides with v outside K and on its periphery, but which within K is defined by the Poisson integral generated by the values of v on its periphery. Condition II may then be written

$$u = M_K u.$$

The following definitions are now made:[4] *Every function ϕ continuous on G and satisfying the following conditions is a lower function of the continuous function f:*

$$\overline{\phi} \leq f \text{ on the boundary } R;$$

$$\phi \leq M_K \phi \text{ for every sphere } K.$$

Every function ϕ continuous on G and satisfying the following conditions is an upper function of f:

$$\underline{\phi} \geq f \text{ on } R;$$

$$\phi \geq M_K \phi \text{ for every } K.$$

Perron then proves that the assemblage of all upper functions has a finite lower bound, and that this lower bound is harmonic in G. Similarly the assemblage of all lower functions has a finite upper bound, and this upper bound will be harmonic. Perron shows that under certain assumptions which we shall not stop to consider, but which are substantially identical with those made by Kellogg in the paper which we have cited, these two functions coincide, and assume the assigned boundary values continuously. Even in the case where the Dirichlet problem is not solvable, these harmonic functions may both be regarded as solutions of a generalized Dirichlet problem. However, there is a disadvantage in this: first, they have not been proved to depend linearly on

[4] Perron demands that an "Oberfunktion" or "Unterfunktion" be continuous on $R+G$. In none of his proofs, however, is anything more than continuity on the interior+boundary of K demanded. The Perron definition will not yield the results here desired without the Perron conventions as to Riemann surfaces and spaces, which we expressly do not wish to introduce.

the boundary conditions, and secondly, they are two functions, each of which has an equal right to be considered as a solution of the generalized Dirichlet problem. The theory is much simplified when we can find a single function to play that rôle.

Now, in one of my papers cited above, I have shown that for any continuous f, the Dirichlet problem admits a certain unique solution in a generalized sense. To be more precise, my theorem reads as follows: *Let G, R, and f be defined as above. Then*

(1) *There exists a sequence $\{R_n\}$ of boundaries such that the Dirichlet problem is solvable in the ordinary sense over the interior of each R_n for all continuous boundary conditions; such that R_n always lies in the interior of R_{n+1}; and such that every point of G is interior to some R_n.*

(2) *There exists a function $v(P)$ continuous over $R+G$, and coinciding with $f(P)$ on R.*

(3) *If $u_n(P)$ is the harmonic function determined on the interior of R_n by the boundary values v on R_n, then the sequence of functions u_n converges at any point P of G to a limit $u(P)$. This limit will be a harmonic function of P throughout G, and will be independent of the choice of the sequence $\{R_n\}$ and the function $v(P)$.*

If we agree that our harmonic functions are to vanish at infinity, an exactly similar theorem will hold for the exterior problem.

What I here wish to prove is that *for any continuous f and for any open-bounded G, u will be at once the lower bound of all upper functions and the upper bound of all lower functions.* That is, the apparent difficulties with the Perron generalization of the Dirichlet problem are not actual, and the generalization completely coincides with that which I have given. In addition, Perron's form of generalization is far more elegant than mine, in that no more or less artificial use is made of such auxiliary elements as $v(P)$ or the sequence $\{R_n\}$. It is furthermore better adapted to the direct study of discontinuous boundary conditions.

As a preliminary, let us discuss the simplest possible case. We have given two bounded closed sets of points Σ_1 and Σ_2 with no point in common. Let us now consider the generalization of the Dirichlet problem corresponding to the boundary values 1 on Σ_1 and 0 on Σ_2. As the region exterior to Σ_1 and Σ_2 will not in general be closed, the Dirichlet problem which we are considering is an

exterior problem, and it is necessary to assume that the desired harmonic function will vanish at infinity. We shall also assume that $v(P)$ is 1 in the neighborhood of Σ_1 and 0 in the neighborhood of Σ_2. Under these conditions the theorem of my previous paper reads as follows: *Let* $\{R_n\}$ *be a sequence of boundaries, each included in its predecessor, containing in common every point of* Σ_1 *and no other point. Furthermore, let these be so chosen that the exterior Dirichlet problem is solvable for each. Let* $\{S_n\}$ *be a sequence of boundaries satisfying exactly the same conditions as those laid down for* $\{R_n\}$, Σ_2 *being substituted for* Σ_1. *Let* $\phi_{m,n}$ *be the harmonic function vanishing at infinity and corresponding to boundary values* 1 *on* R_m and 0 *on* S_n. $\phi_{m,n}(P)$ *will always exist for sufficiently large values of* m *and* n, *and as* m *and* n *increase to infinity in any way,*

$$\lim \phi_{m,n}(P) = u(P)$$

exists for all P*'s not within* Σ_1 *nor* Σ_2. *Furthermore,* u *is harmonic everywhere outside* Σ_1 *and* Σ_2.

Let it be noted that $\phi_{m,n}(P)$ is a monotone non-decreasing function of n and a monotone non-increasing function of m. It is also bounded. Hence

$$\lim_{n\to\infty} \phi_{m,n}(P) = \phi_m(P)$$

and

$$\lim_{m\to\infty} \phi_{m,n}(P) = \phi^{(n)}(P)$$

exist. We shall make the convention that within R_m, $\phi_m(P) = 1$, and within S_n, $\phi^{(n)}(P) = 0$. There is no difficulty in showing that

$$\lim_{m\to\infty} \phi_m(P) = \lim_{m\to\infty} \phi^{(m)}(P) = u(P)$$

everywhere outside Σ_1 and Σ_2.

I now wish to show that ϕ_m is an upper function and that $\phi^{(m)}$ is a lower function for boundary conditions 1 on Σ_1 and 0 on Σ_2. To begin with, ϕ_m assumes limiting values on Σ_2 which are never negative and is 1 on Σ_1. Hence ϕ_m is never less than the boundary values. In the second place, if K is entirely outside R_m, $\phi_m = M_K\phi_m$. If K cuts R_m, $M_K\ \phi_m$ cannot exceed 1 on the part of R_m

interior to K. Hence over the contour constituted by this part of R_m and the part of K exterior to R_m,

$$\phi_m \geq M_K \phi_m.$$

Therefore this relation must hold within this contour, as the two functions are harmonic. Over the rest of the interior of K,

$$1 = \phi_m \geq M_K \phi_m.$$

That is, ϕ_m is an upper function. The argument in the case of $\phi^{(m)}$ is precisely parallel. Since, however, u is the common limit of the sequences $\{\phi_m\}$ and $\{\phi^{(m)}\}$, and since no lower function anywhere exceeds an upper function, the upper bound of all lower functions coincides with the lower bound of all upper functions, and is u itself.

We may immediately proceed from this to the case where our boundary consists of a finite number of closed, bounded, disconnected pieces, while f assumes a constant value over each piece. Let the pieces be $\Sigma_1, \ldots, \Sigma_n$, and let $f = a_j$ on Σ_j. Let a be the least of the a_j's, and let $f_j = a_j - a$ on Σ_j and $f_j = 0$ on the other Σ_k's. Then

$$f = \Sigma f_j + a.$$

Furthermore, if g_j is an upper function for f_j, then

$$g = \Sigma g_j + a$$

is an upper function for f. Similarly, if h_j is a lower function for f_j, then

$$h = \Sigma h_j + a$$

is a lower function for f. We have, however, proved in substance that g_j and h_j can be chosen as close as we like to one another. Hence we can make the difference between g and h as small as we like, so that the lower bound of all upper functions of f will be at the same time the upper bound of all lower functions of f. Furthermore, as I have shown in a previous paper that my solution of the generalized Dirichlet problem depends linearly on the boundary conditions, it follows, as a moment's consideration will show, that the lower bound of upper functions of f and the upper bound of lower functions of f assume this solution as their common value.

It is now necessary to make the transition to the most general case of continuous boundary conditions. For this purpose, I must recall some definitions and theorems from my previous paper. At a point P the *relative weight* of a point-set Σ_1 to a point-set Σ_2 is the potential corresponding in the sense already explained to a potential 1 on the set consisting of Σ_1 together with its derivative, a potential 0 on the set consisting of Σ_2 together with its derivative, and a potential 0 at infinity. This definition is only directly applicable in case Σ_1 and Σ_2 have no point nor limit-point in common. The *outer weight* at P of a part σ of a bounded point-set Σ is the lower bound of the relative weight of σ to point-sets τ in Σ containing no point nor limit-point in common with σ. It is easy to show that if σ and σ' are two mutually exclusive closed sub-sets of Σ, the outer weight of $\sigma+\sigma'$ is precisely the sum of the outer weights of σ and of σ'. Furthermore, if σ_1 includes σ, the outer weight of σ_1 is at least that of σ.

Now let Σ_ξ consist of those points of Σ with an X-coördinate not exceeding ξ. Considered as a function of ξ, Σ_ξ is monotonic and nondecreasing. It has hence only a denumerable set of discontinuities. If Σ_ξ is continuous for $\xi=\xi_1$, it follows that if ϵ is any positive number, we can choose a positive η so small that

$$\text{outer weight of } [\Sigma_{\xi_1+\eta}-\Sigma_{\xi_1-\eta}]<\epsilon.$$

Hence if σ consists of all the points of Σ with X-coördinates between $\xi_1-\eta/2$ and $\xi_1+\eta/2$, inclusive, it follows from the theorem which we have just proved that

$$\text{outer weight of } \sigma<\epsilon.$$

It consequently follows that for all but a denumerable set of values of ξ, the plane $x=\xi$ may be surrounded by a region of arbitrarily small outer weight. Now let ξ_1 and ξ_2 be the extreme X-coördinates of Σ. The interval (ξ_1, ξ_2) may be subdivided into a finite number of sub-intervals of length less than $\theta/2$ in such a way that no point of subdivision is one of the denumerable set of points of discontinuity of Σ_ξ. Hence the entire set of points of Σ contained in the planes perpendicular to the X-axis and passing through the points of subdivision of the interval (ξ_1, ξ_2) on the X-axis may be surrounded by a region of outer weight

less than ϵ. *Mutatis mutandis*, the same thing applies with regard to the Y and Z axes. In other words, given any positive number θ, the closed set Σ may be subdivided into a finite number of mutually exclusive closed compartments with greatest dimension not exceeding θ, plus a set of points contained in three sets each of outer measure less than ϵ, and consequently of outer measure not exceeding 3ϵ.

Let us now consider a continuous function $f(P)$ defined over Σ. Let θ be so chosen that the oscillation of $f(P)$ in a closed region of maximum dimension never exceeds ϵ. Let us form an upper function and a lower function to f in the following manner.

(1) We form a set of compartments in the manner indicated in the last paragraph; we let $\psi_1(P)$ be an upper function of the step-function assuming in each compartment as its value the upper bound of f in that compartment; and we let $\psi_2(P)$ be a lower function of the step-function assuming in each compartment as its value the lower bound of f in that compartment. It follows from what we have said that if χ_1 and χ_2 are respectively these step-functions, we may determine ψ_1 and ψ_1', upper and lower functions to χ_1 such that for a given, P in G

$$\psi_1(P)-\psi_1'(P)<\epsilon.$$

We can similarly determine ψ_2' and ψ_2, upper and lower functions respectively to χ_2, such that for this P

$$\psi_2'(P)-\psi_2(P)<\epsilon.$$

These upper and lower functions may indeed be formed as polynomials in the upper (or lower) functions to the different functions assuming only the values 0 and 1 from which χ_1 and χ_2 are built up as sums. We see at once that

$$\psi_1(P)-\psi_2(P)<3\epsilon.$$

(2) We now consider τ, the set of all points of Σ which are not contained in the compartments of (1). Since the outer weight of τ with respect to Σ does not exceed 3ϵ, we may determine a function $\psi_3(P)$, harmonic except on Σ and within a certain small distance of points of τ, continuous except on Σ, non-negative, equalling 1 in the given neighborhood of τ, and not exceeding 4ϵ at the given P.

(3) We write

$$\Psi_1(P)=\psi_1(P)+\max|f|.\psi_3(P);$$
$$\Psi_2(P)=\psi_2(P)-\max|f|.\psi_3(P).$$

Ψ_1 is clearly an upper function to f, and Ψ_2 a lower function. Furthermore,

$$\Psi_1(P)-\Psi_2(P)<\epsilon(3+8\max|f|).$$

As ϵ is arbitrary, this may be made as small as we like. Hence the upper bound of all lower functions coincides with the lower bound of all upper functions.

(4) In my previous paper, I proved that the generalized solution which I there obtained in the general case was the limit of the harmonic function determined by the boundary values χ_1. This function must lie between ψ_1 and ψ_1', and hence between ψ_1 and ψ_2. *A fortiori* it lies between Ψ_1 and Ψ_2. Hence the function which constitutes the solution of the generalized Dirichlet problem of my previous paper is the upper bound of all lower functions of its boundary values and the lower bound of all upper functions.

§2. **Discontinuous Boundary Values.** In the simplest cases of discontinuous boundary values, the Perron generalization of the Dirichlet problem is still satisfactory. For example, let our boundary function be 0 except at a single point of the boundary, where it is 1. It is then possible at a point P interior to the boundary to find an upper function $\phi(P)$ approximating to 0 as closely as we wish. Since, however, 0 is itself a lower function, the upper bound of all lower functions is at the same time the lower bound of all upper functions.

However, we do not need to go much further to find an instance where the Perron generalization breaks down. For example, let G be the interior of a circle with boundary R. Let us establish a system of polar coördinates with the center of G as pole. Let $f(P)$ be 1 for those points P of R with a θ-coördinate rational in terms of 2π, and let $f(P)$ be 0 elsewhere on R. Clearly any upper function of f is at least 1, and no lower function is greater than 0. Hence the Perron method yields us here no unique generalization of the Dirichlet problem.

The difficulty here is of the same character as the difficulty in applying an ordinary integral to the boundary values in question. In this case the Riemann definition breaks down, and we must have recourse to the definition of Lebesgue. In the case of the Dirichlet problem, then, we must develop an analogue to the Lebesgue integral. Now, the fundamental property of the Lebesgue integral is that if $f_1, f_2, \ldots$ are all uniformly bounded and integrable over a closed set of points S, and if $f = \lim f_n$ exists for every point in S, then f is also integrable, and in fact its integral is the limit of the integrals of the f_n's. This suggests the following definitions, which are confined to the case of bounded functions f: *if* $f(P)$ *is a function defined on* R; *if* $f(P) \leq \lim_{n \to \infty} f_n(P)$ $[f(P) \geq \lim_{n \to \infty} f_n(P)]$ *everywhere on* R; *if every* f_n *is continuous, and the set of all* f_n*'s is bounded; if* $\phi_n(P)$ *is an upper [lower] function of* $f_n(P)$; *and if* $\phi_n(P)$ *tends to* $\phi(P)$ *with increasing* n *at every point* P *of* G; *then* $\phi(P)$ *will be called a superfunction [subfunction] to* f.

We now wish to do two things: first to prove that every superfunction of f is at least as large as any subfunction of f; and second to discuss those cases in which the lower bound of all superfunctions coincides with the upper bound of all subfunctions. The common value of these two bounds (which will always be harmonic functions) will be regarded as a solution of the generalized Dirichlet problem corresponding to boundary values f.

Both of our ends can be accomplished with the aid of the Daniell integral and of the methods of a paper of mine entitled *Discontinuous Boundary Conditions and the Dirichlet Problem*. I there pointed out that the potential at a point P of G corresponding to boundary values f on R may be regarded as a type of Daniell integral of f, and in my later papers, I indicated how this result could be extended to cases where the Dirichlet problem was only solvable in the generalized sense. Without going into the assumptions necessary for the application of the Daniell theory, which are fully discussed in my previous paper in so far as they apply to the case in question, the outline of the Daniell theory is the following. We start with a class of functions T_o on a general assemblage, and an operation I on the functions of the class. T_o is here the class of all functions f continuous on R, and $I\{f\}$

is the potential corresponding (in my generalized sense) to the set of boundary values f. We then form the class T_1 of functions obtained as limits of monotone non-descending sequences of functions of T_0. For these functions, the result of applying the operator I is defined as the limit, finite or infinite, of the I's of the functions of the sequence. This definition is always applicable and unambiguous. We next define the upper integral of a function f as the lower bound of the I's of functions of T_1 nowhere less than f, and the lower integral of f as the negative of the upper integral of $-f$. If the upper and lower integrals of f coincide and are finite, f is said to be summable, and the common value of the upper and lower integrals is $I\{f\}$. Daniell proves that if the conditions for the validity of his theory are satisfied, the different stages of the definition of I are compatible; that if f_n tends to f in such a way that there is a summable function exceeding in absolute value every f_n, while every f_n is summable, then f_n is summable, and in fact

$$I\{f\} = \lim I\{f_n\};$$

that if $f \geq g$ for every argument, $I\{f\} \geq I\{g\}$; that I is linear; that the class of summable functions is linear, and contains the absolute value of any one of its members.

It follows from Daniell's definitions and theorems that if we write $I_P\{\psi\}$ for the value at P of the harmonic function determined in accordance with the Daniell generalization process by the boundary values ψ, then if we employ the notation which we used in the definition of superfunctions,

$$\phi(P) = \lim \phi_n(P) \geq \lim I_P\{f_n\} \geq \bar{I}_P\{f\},$$

where the symbol $\bar{I}$ is used to represent the upper integral. Similarly, if ϕ_1 is a subfunction of f, we shall have

$$\phi_1(P) \leq \underline{I}_P\{f\},$$

where $\underline{I}_P\{f\}$ stands of course for the lower Daniell integral of f. Since Daniell proves that

$$\bar{I}_P\{f\} \geq \underline{I}_P\{f\},$$

we see that any superfunction is at least as large as any subfunction. Furthermore, if the upper bound of all subfunctions coincides

with the lower bound of all superfunctions, the common value must also be the common value of $\bar{I}_P\{f\}$ and $\underline{I}_P\{f\}$, and must hence be $I_P\{f\}$.

If $f(x)$ is a bounded member of T_1, the upper bound of all its subfunctions and the lower bound of all its superfunctions must coincide. To show this, it is only necessary to point out that we can represent f as $\lim f_n$, where $\{f_n\}$ is an increasing, bounded sequence of continuous functions. We may accordingly choose an upper function $\phi_1(P)$ of $f_1(P)$ less than $\epsilon+$ the lower bound of all upper functions of f_1, an upper function $\phi_2(P)$ of $f_2(P)$ less than $\epsilon/2+$ the lower bound of all upper functions of f_2, and so on. We choose $\overline{\phi}_n(P)$ as a lower function of f_n greater than

$$(\text{upper bound of lower functions of } f_n) - \epsilon/n.$$

Then since $\{I_P\{f_n\}\}$ is a monotonic and hence a convergent sequence, and since $I_P\{f_n\}$ is at once the lower bound of all upper functions and the upper bound of all lower functions of f_n, we have

$$I_P\{f\} = \lim I_P\{f_n\} = \lim \phi_n(P) = \lim \overline{\phi}_n(P).$$

Thus $I_P\{f_n\}$ is at once a superfunction and a subfunction of f.

The transition from T_1 to all bounded summable functions is easy, since it can be shown that any such function may be penned in between two *bounded* functions, one belonging to T_1 and t other the negative of a function belonging to T_1, with " integrals " differing by less than ϵ. Any superfunction to the larger of these two functions is a superfunction to the summable function between, while any subfunction to the smaller of the two confining functions is a subfunction to the function between. From this it may be concluded at once that for any bounded summable function on G, the upper bound of its subfunctions, the lower bound of its superfunctions, and the solution of the generalized Dirichlet problem in the sense of my previous paper, all coincide.

The condition that the summable function be bounded may be eliminated if in the definition of superfunctions and subfunctions, we substitute for the boundedness of the set of f_n's the following condition: *Let $F_n(P)$ be the function always assuming the greatest value of the quantities $|f_1(P)|, \ldots, |f_n(P)|$. Let the set of quantities $I_P\{F_n\}$ be bounded.*

THE TORSION PROBLEM OF CURVED BEAMS*

By PAUL HEYMANS† AND W. J. HEYMANS‡

Part 1

Given a beam in elastic equilibrium whose mean fibre is a continuous curve and which is acted upon by couples of arbitrarily directed axes, the problem which we set is to determine:

I. *The system of forces acting on any section normal to the mean fibre.*

II. *The stresses existing in this section and the limiting of those stresses.*

I. Determination of the forces acting on a section.

We first consider a section ABC of a plane curved beam represented in Fig. 1 by its mean fibre, with a torsional moment applied

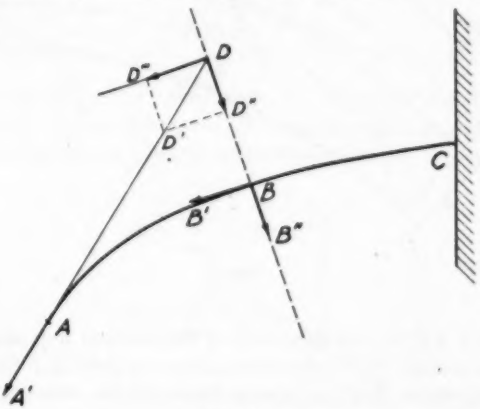


Fig. 1

at A in a plane normal to the mean fibre. The moment applied at A is represented by the vector $\overline{AA'}$ following the ordinary

*Read by title at the meeting of the National Academy of Sciences, Nov. 10, 1924. Abstract in The Proc. Nat. Acad. Sc., Jan. 1925.

†Massachusetts Institute of Technology, Cambridge, Mass., U. S. A.

‡University of Ghent, Ghent, Belgium.

conventions of mechanics. To determine the forces acting on a section B normal to the mean fibre at B, we prolong $\overline{AA'}$ until it intersects the trace BD of the section. At D we carry $\overline{DD'} = \overline{AA'}$ which we decompose, in conformity with the de Saint Venant principle of elastic equivalence of statically equivalent systems, into its components DD''' and DD'', tangential and normal respectively to the trace BD. Then the vectors $\overline{BB'}$ and $\overline{BB''}$ brought into B, equal to $\overline{DD'''}$ and $\overline{DD''}$ are the vectors representative of the moments acting on the section B.

In the case of a bar of rectangular cross section, Fig. 2, the

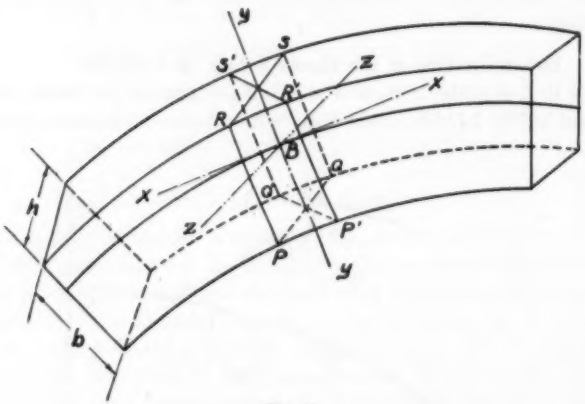


Fig. 2

vector $\overline{BB'}$ will be representative of the moment imposing torsion, upon the section $PQRS$, about an axis xx tangent at B to the mean fibre; the vector $\overline{BB''}$ is representative of the moment imposing simple flexion, upon the same section, about the neutral axis yy, $P'Q'R'S'$ being the position taken by the section $PQRS$ under the action of the bending moment.

The above known method can be readily extended to a beam whose mean fibre is any continuous curve which, for instance, projects in $ABCOD$ in Fig. 3. Let this beam be acted upon by moments represented by their respective vectors which project

in AA', BB', etc. If we consider at O a normal section to the mean fibre and m_k' and l_k' be the cosines of the angles made by

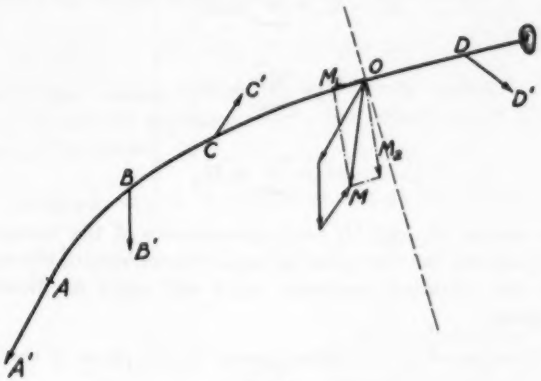


Fig. 3

the k vectors $\overline{AA}'$, $\overline{BB}'$, etc., with the plane of the section O and a normal to this plane, the forces acting upon this section O may be represented by the two moment-vectors M_1, M_2, where

$$\left.\begin{aligned} M_1 &= \sum_1^k l_k' M_k \\ M_2 &= \sum_1^k m_k' M_k \end{aligned}\right\} \qquad (1)$$

The summation extends over the k moments acting on the isolated section $ABCO$.

The vector M_1 is representative of the moment imposing torsion upon the section about an axis tangent to the mean fibre. The vector M_2 is representative of a moment of bending around an axis parallel to the vector M_2. If the cross section O admits two principal axes of symmetry y and z and the angular directions of the vectors M_k withthe directions of y and z and a normal to

the section be respectively m_k, n_k, l_k, equations (1) will take the form:

$$\left.\begin{aligned} M_1 = M_x &= \sum_1^k l_k M_k \\ M_y &= \sum_1^k m_k M_k \\ M_z &= \sum_1^k n_k M_k \end{aligned}\right\}. \qquad (2)^*$$

The vectors M_y and M_z are representative of the moments of bending around the two principal axes of symmetry of the section.

The two following particular cases will serve as illustrative verifications:

1. Consider, Fig. 4, a beam curved in the plane of the figure

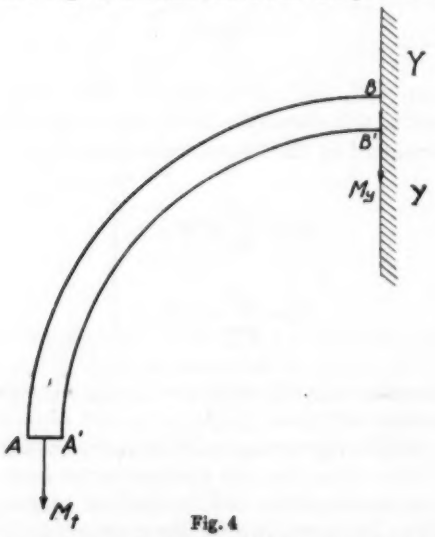


Fig. 4

*The moment vectors M_x, M_y, M_z could also be obtained by graphical addition of the vectors $\overline{AA'}$, $\overline{BB'}$ etc. and subsequent decomposition of the resultant vector into the x, y and z components.

into a quarter circle and leaning by its plane end BB' against the plane yy. Let M_t be a torsional couple applied at AA' in the plane normal to the mean fibre. We shall obtain for the section BB':

$$M_x = M_z = O \qquad M_y = M_t.$$

That is: the section BB' of the beam will undergo a simple flexion around the neutral axis yy. The correctness of the result is readily visualized.

2. Consider, Fig. 5, a cylindrical bar curved into a quarter

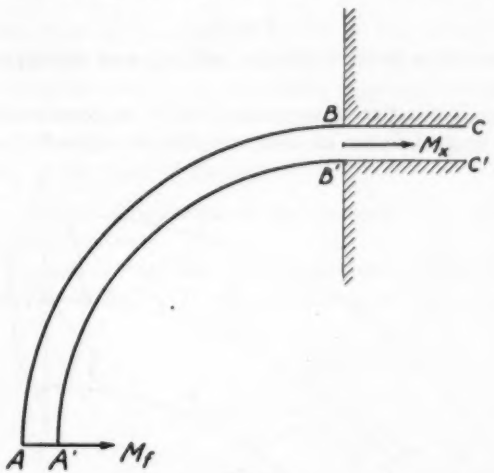


Fig. 5

circle from AA' to BB' and straight from BB' to CC', the straight part being engaged in a straight cylindrical sleeve. We apply at AA' a flexional moment represented by the vector moment M_f.

We shall obtain for BB':

$$M_y = M_z = O \qquad M_x = M_f.$$

That is: the portion from BB' to CC' undergoes a torsional action which will result in a smooth counterclockwise rotation of the bar around the axis xx, parallel to the mean fibre.

II. **Calculation of the Stress-Components.**

The stress components due to each of the moments M_x, M_y and M_z can now be calculated separately and then superposed. We shall see that *we thereby arrive at a solution of the problem of the stresses in plane and skew beams acted upon by couples for all cases where separate solutions exist for the problem of torsion of the straight beams and of flexion of the plane curved beams.* We postulate, however, the validity of the principle of superposition of forces and their effects. The limitation of the stresses will also be considered in the second part of the paper.

Part 2

Application to bars of circular, elliptic, and rectangular cross sections.

We consider a bar whose mean fibre is any continuous curve. Let the applied forces in a section AB be reduced to a couple

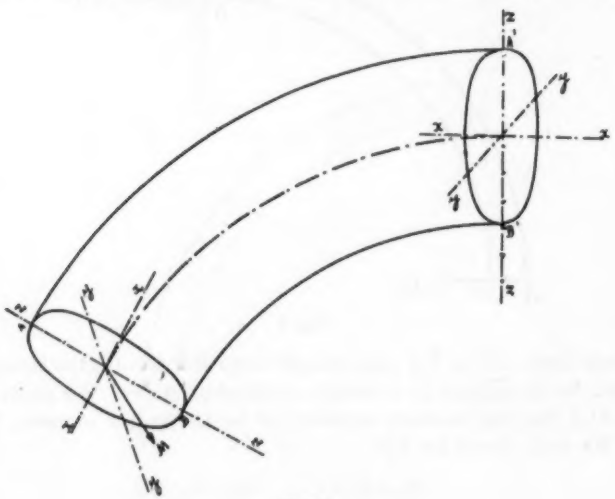


Fig. 6

represented (Fig. 6) by the moment vector M, the direction cosines of this vector with respect to the three axes x, y, z, being

l, m, n. Projecting this vector-moment M on the three axes, the three couples M_x, M_y, M_z are obtained, where

$$M_x = lM, \qquad M_y = mM, \qquad M_z = nM$$

The moment represented by M_x will cause torsion, whereas the moments represented by M_y and M_z will cause bending respectively around the y axis and the z axis. In order to obtain the applied forces in any other section $A'B'$ due to the moment M, we prolong the vector M until its intersection with the plane $A'B'$, transport this vector parallel to itself on the mean fibre in $A'B'$ and proceed to the same decomposition as in the section AB.

The direction of the moment vector M has been arbitrarily chosen and it is thereby seen that the case which we consider introduces all the other types of outside applied forces which can be reduced to couples. If it is desired to consider solely the torsion of curved bars, it will suffice to take the original vector M perpendicular to the plane of the section AB.

Calculation of the Stresses[2]

1. Circular Section.

A. Torsional Stresses due to the Component Couple M_x.

The section AB (Fig. 7) will rotate around the x axis with

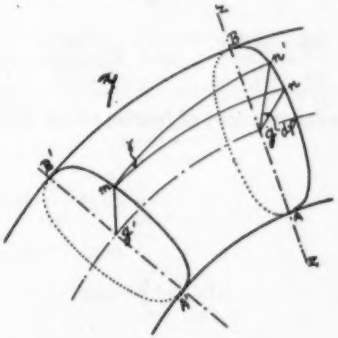


Fig. 7

respect to an infinitely near section $A'B'$: a fibre such as mn will

[2] The notations for the stresses and strains are those of A. E. H. Love, Treatise on the Mathematical Theory of Elasticity.

be displaced into $m'n'$. Following the ordinary theory, if $d\phi$ is the angle which the radius gn' makes with the radius gn, we have, taking $mn=ds$:

$$nn'=rd\phi=\gamma ds, \quad \text{where } \gamma=r\frac{d\phi}{ds}$$

and the torsional stress $X_t=G\gamma=Gr\frac{d\phi}{ds}$ where G is the torsional modulus of elasticity. These relations have been written for a point at the boundary. They hold for any point at distance r from the origin.

The value of $\frac{d\phi}{ds}$ is obtained from the following equation of equilibrium of rotation of the section AB.

$$\int_\Omega X_t r d\omega = M_z$$

where the integration extends over the area Ω of the cross section.

Hence

$$M_z=G\frac{d\phi}{ds}\int_\Omega r^2 d\omega$$

or:

$$\frac{d\phi}{ds}=\frac{M_z}{GI'} \qquad \text{and } X_t=\frac{M_z r}{I'}$$

where I' is the polar moment of inertia of the section with respect to the center.

B. Bending around the y axis due to the component couple M_y. The couple M_y will cause a rotation of the section AB with respect to $A'B'$, around the y axis which projects in g' (Fig. 8).

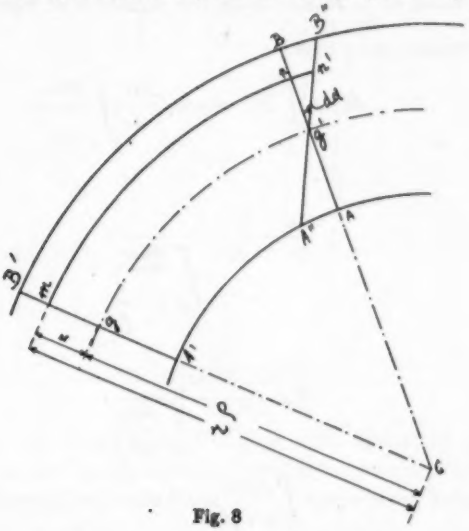


Fig. 8

The normals to the mean fibre at the points g and g' intersect at the center of curvature C.

The dilatation e_{xx} of a fibre mn will be given by:

$$e_{xx} = \frac{nn'}{mn}$$

where $nn' = z\,da$ and $\frac{mn}{gg'} = 1 + \frac{z}{\rho}$.

Writing $gg' = ds$, we obtain:

$$mn = \left(1 + \frac{z}{\rho}\right) ds$$

and

$$X_x' = E e_{xx} = E \frac{z \frac{da}{ds}}{1 + \frac{z}{\rho}}$$

The stress X_x' is normal to the section which is tangential to the fibre mn and is thus parallel to the x axis.

The value of $\frac{da}{ds}$ results from the equation of equilibrium of rotation around the y axis:

$$M_y = \int_\Omega X_x' \, z d\omega = E \frac{da}{ds} \int_\Omega \frac{z^2 d\omega}{1+\frac{z}{\rho}}$$

or:

$$\frac{da}{ds} = \frac{M_y}{E} \cdot \frac{1}{\int_\Omega \frac{z^2 d\omega}{1+\frac{z}{\rho}}}.$$

Hence

$$X_x' = \frac{M_y}{(z+\rho)\int_\Omega \frac{z^2 d\omega}{z+\rho}}.$$

The value of the integral $\int_\Omega \frac{z^2 d\omega}{z+\rho}$ is calculated in Appendix 1 where it is found that:

$$\int_\Omega \frac{z^2 d\omega}{z+\rho} = 2\pi\rho\left(\rho^2+a^2\right)\left(\frac{\rho}{\sqrt{\rho^2-a^2}}-1\right)-\pi a^2\rho$$

ρ being the radius of curvature of the bar. We replace the value of the integral by K_1.

It is seen that the stress X_x' follows a hyperbolic variation. Indeed for $z=\infty$, X_x' becomes:

$$X_x' = \frac{M_y}{\left(1+\frac{\rho}{z}\right)K_1} = \frac{M_y}{K_1}$$

the stress changing sign with the moment M_y. For $z=-\rho$ the stress X_x' equals infinity. The asymptotes of the hyperbola are thereby known.

Moreover, for $z=0$, the stress X_x' becomes zero. Consequently, according to the sign of M_y, the values of X_x' will be given by

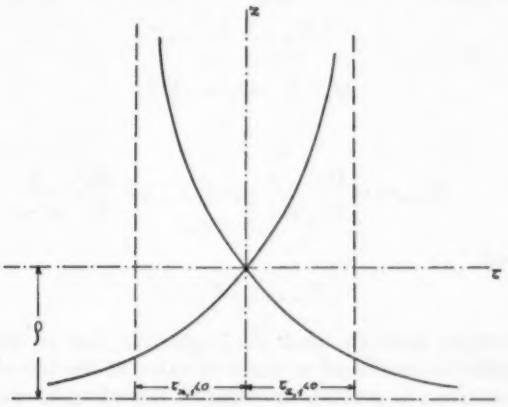


Fig. 9

one of the two hyperbolæ of Fig. 9. The points of maximum stress lie on the outside fibres, namely in b or d (Fig. 10).

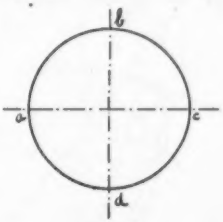


Fig. 10

Indeed, if $M_y>0$, *i.e.*, if the couple tends to increase the curvature of the bar, supposed to be positive, the stresses $X'_{x,b}$ and $X'_{x,d}$ at the points b and d are respectively:

$$X'_{x,b}=\frac{M_y}{K_1}\frac{a}{\rho+a} \text{ and } X'_{x,d}=-\frac{M_y}{K_1}\frac{a}{\rho-a}$$

where a is the distance $g'B$ or $g'A$ (Fig. 8).

As

$$\rho > a,\ (\rho - a) < (\rho + a)$$

$$|X'_{x,d}| > |X'_{x,b}|.$$

If

$$M_y < 0,\quad M_y = -M_y',$$

then:

$$X'_{x,b} = -\frac{M_y'}{K_1}\ \frac{a}{\rho + a} \text{ and } X'_{x,d} = \frac{M_y'}{K_1}\ \frac{a}{\rho - a}$$

and again

$$|X'_{x,d}| > |X'_{x,b}|.$$

We therefore conclude that the stress X_x' due to the couple M_y reaches its maximum in absolute value at the intrados fibre. This maximum is positive or negative according to the sign of M_y.

C. Bending around the z axis due to the component couple M_z. Under the action of the couple represented by the moment

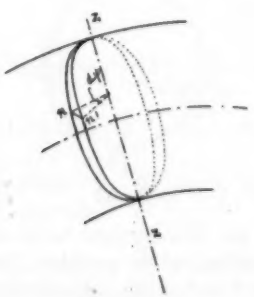


Fig. 11

vector M_z, the section will rotate around the z axis so that points such as n will come in n' (Fig. 11).

Referring to Fig. 12, the dilatation will be given by:

$$e_{xx}=\frac{nn'}{mn} \qquad \text{(Fig. 12)}$$

where

$$nn'=yd\psi \qquad \text{(Fig. 11)}$$

and

$$\frac{mn}{gg'}=\frac{z+\rho}{\rho}$$

or:

$$mn=\left(1+\frac{z}{\rho}\right)ds$$

Hence:

$$e_{xx}=\frac{y\dfrac{d\psi}{ds}}{1+\dfrac{z}{\rho}}.$$

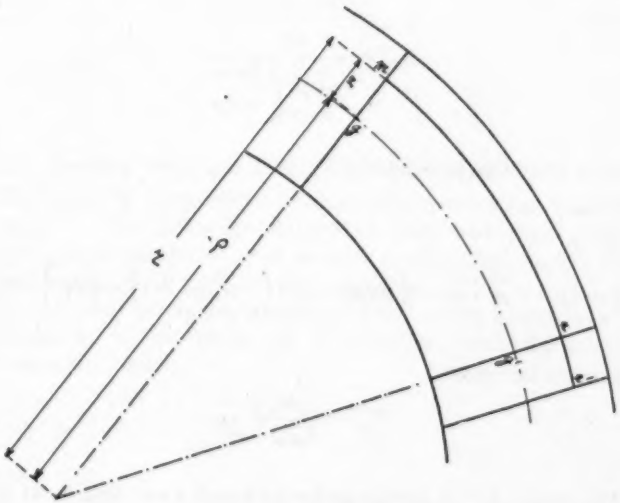


Fig. 12

The corresponding stress is directed tangentially to the fibre mn, which is parallel to the x axis. We therefore have:

$$X_x'' = E \frac{y \frac{d\psi}{ds}}{1+\frac{z}{\rho}}.$$

Writing the equation of equilibrium of rotation around the z axis, we obtain:

$$\int_\Omega X_x'' y \, d\omega = M_z$$

or:

$$\frac{d\psi}{ds} = \frac{M_z}{E} \frac{1}{\int_\Omega \frac{y^2 d\omega}{1+\frac{z}{\rho}}}.$$

Hence:

$$X_x'' = \frac{M_y}{(z+\rho)\int_\Omega \frac{y^2 d\omega}{z+\rho}}.$$

Having according to Appendix 2:

$$\int_\Omega \frac{y^2 d\omega}{\rho+z}$$

$$= \tfrac{2}{3}\pi a^2(\rho - \sqrt{\rho^2-a^2}) - \tfrac{1}{3}\left[2\pi\rho(\rho^2+a^2)\left(\frac{\rho}{\sqrt{\rho^2-a^2}}-1\right) - \pi a^2\rho\right] = K_2$$

The stress becomes:

$$X_x'' = \frac{M_z y}{K_2(\rho+z)}.$$

The stress X_x'' is maximum for $z=0$ and $y=a$; that is, at the points a and c of Fig. 10.

2. Elliptic section.

Directing the z axis along the major axis of the ellipse, the equation of the elliptic cross section (Fig. 13) is:

$$\frac{z^2}{a^2}+\frac{y^2}{b^2}=1.$$

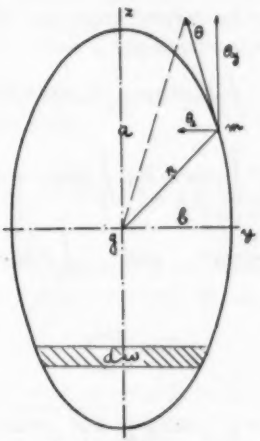


Fig. 13

A. Torsional Stress due to the Component Couple M_x.

The theory of Coulomb for circular cross sections can no more be applied. The following method of Bach and Föppl gives, in the case of an elliptic cross section, results which satisfy the general equations of the theory of elasticity.

At any point m on the boundary of the ellipse having as coordinates x_1, y_1, the stress X_t is tangential to the boundary. We therefore have:

$$\frac{X_{z_1}}{X_{y_1}}=\frac{dz_1}{dy_1}=-\frac{a^2y_1}{b^2z_1}$$

where

$$X_t=\pm\sqrt{X_{z_1}^2+X_{y_1}^2}.$$

The sign of the moment of X_t about the x axis will remain the same all along the periphery of the ellipse.

Admitting that the law of variation of X_t is linear along the radius vector gm, the stress components at any point n can be written:

$$X_x = -Ka^2y, \qquad X_y = Kb^2z.$$

The value of K can be derived from the following equation of equilibrium of rotation around the x axis:

$$\int_\Omega X_y z d\omega - \int_\Omega X_z y d\omega = M_x$$

or:

$$Kb^2 \int_\Omega z^2 d\omega + Ka^2 \int_\Omega y^2 d\omega = -M_x.$$

As

$$\int_\Omega z^2 d\omega = \tfrac{1}{4}\pi a^3 b \quad \text{and} \quad \int_\Omega y^2 d\omega = \tfrac{1}{4}\pi a b^3.$$

Hence

$$K = \frac{2M_x}{\pi a^3 b^3}$$

and:

$$X_t = \frac{2M_x}{\pi a^2 b^2} \sqrt{\frac{a^2y^2}{b^2} + \frac{b^2z^2}{a^2}}$$

At the periphery we have:

$$y^2 = b^2\left(1 - \frac{z^2}{a^2}\right),$$

consequently:

$$X_t = \frac{2M_x}{\pi a^2 b^2} \sqrt{a^2 - \frac{z^2}{a^2}(a^2 - b^2)}$$

This value of X_t is maximum for $z = 0$, namely at the ends of the minor axis, and:

$$(X_t)_{\max} = -\frac{2M_x}{\pi a b^2}$$

The minimum is obtained for $z = a$ and:

$$(X_t)_{\min} = \frac{2M_x}{\pi a^2 b}.$$

B. Bending around the y axis due to the Component Couple M_y.

The following equation has been derived for the case of a circular cross section irrespective of the nature of the cross section. We therefore have:

$$X_x' = \frac{M_y z}{(\rho+z)\int_\Omega \frac{z^2 d\omega}{(\rho+z)}} = \frac{M_y z}{K_1(\rho+z)}$$

where the value K_1 of the integral, determined in Appendix 3, becomes for an elliptic section:

$$K_1 = \int \frac{z^2 d\omega}{r} = \left[\frac{2\pi b\rho}{a}\left(a^2+\rho^2\right)\left(\frac{\rho}{\sqrt{\rho^2-a^2}}-1\right)\right] - \pi ab\rho.$$

As for the circular cross section, it can readily be shown by like derivations that the maximum stress in absolute value occurs at the intrados fibre.

C. Bending around the z axis due to the Component Couple M_z.

The following equation has also been derived irrespective of the nature of the cross section. We therefore have:

$$X_x'' = \frac{M_z y}{(\rho+z)\int_\Omega \frac{y^2 d\omega}{(\rho+z)}} = \frac{M_z y}{K_2(\rho+z)}$$

where the value K_2 or the integral, determined in Appendix 3, is:

$$K_2 = \int_\Omega \frac{y^2 d\omega}{\rho+z}$$

$$= \frac{16b^3}{a^3}\left\{\pi a^2\left(\rho-\sqrt{\rho^2-a^2}\right) - \left(\rho^2+a^2\right)\left(\frac{\pi\rho^2}{\sqrt{\rho^2-a^2}}-\pi\rho\right)+\frac{\pi a^2\rho}{2}\right\}$$

This stress is maximum for $z=0$ and $y=\pm b$, namely at the points a and c.

3. Rectangular section.

Let the z axis be directed parallel to the greater side of the rectangle (Fig. 14).

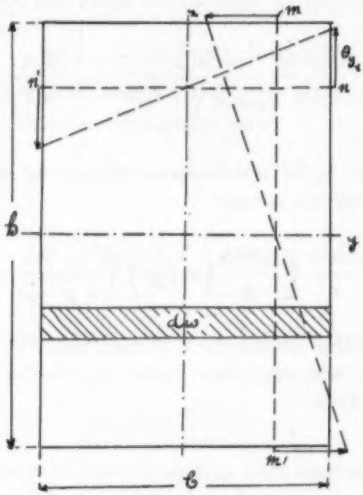


Fig. 14

A. Torsional stress due to the Component Couple M_x.

It is known that the method of Bach, which we shall follow hereunder, yields, in the case of a rectangular cross section, only an approximate solution. Indeed, the stresses thereby obtained do not verify the general equations of the theory of elasticity.

Along the periphery of the cross section, the tangential stress must be at all points tangent to said periphery, and must therefore vanish at the corners of the rectangular cross section. Moreover, the section has two axes of symmetry and the shearing stresses pass necessarily through a maximum on those axes. A parabolic law of variation ascribed to said tangential stress along the periphery will satisfy those conditions.

The law of variation along the side of the rectangle of the X_{y_1} stress existing at m can be written as follows:

$$X_{y_1} = K\left(\frac{b^2}{4} - y^2\right).$$

At a symmetrical point m' the same stress will exist, but of opposite sign.

Likewise at n and n' we will have a stress X_{z_1} whose law of variation along the periphery is given by:

$$X_{z_1} = K'\left(\frac{h^2}{4} - z^2\right).$$

Those laws of variation have been experimentally verified by Bach who has also shown that the axes of symmetry do not undergo any deformation. This justifies the linear variation of X_y and X_z respectively along mm' and nn'. Their values therefore can be expressed as follows:

$$\frac{X_y}{X_{y_1}} = \frac{z}{\frac{h}{2}} \qquad \text{whence } X_y = \frac{2z}{b} K\left(\frac{b^2}{4} - y^2\right)$$

and

$$\frac{X_z}{X_{z_1}} = \frac{y}{\frac{b}{2}} \qquad \text{whence } X_z = -\frac{2y}{b} K'\left(\frac{h^2}{4} - z^2\right).$$

The values of K and K' can, as before, be derived from the equation of equilibrium of rotation around the x axis together with the equation of elastic equilibrium. The following values for the stress components are thereby obtained:

$$X_z = \frac{36 M_x}{b^3 h^3} z\left(\frac{b^2}{4} - y^2\right)$$

$$X_y = -\frac{36 M_x}{b^3 h^3} y\left(\frac{h^2}{4} - z^2\right).$$

In order to obtain the maximum of those stress components, the first derivatives with respect to y and z should be equated to zero and the roots of those equations determined. This leads in many cases to intricate calculations. In special cases it is possible to verify that the analytical maximum, which occurs at a

point lying inside of the section, is smaller than the absolute value of the maximum stress existing at the periphery for $z=0$; namely, on the minor axis. This maximum tangential stress has the following value:

$$(X_t)_{\max} = \frac{9M_x}{2b^2h}.$$

B. Bending around the y axis due to the Component Couple M_y. The following equation has been derived irrespective of the nature of the cross section. We therefore have:

$$X_x' = \frac{M_y z}{(\rho+z)\int_\Omega \frac{z^2 d\omega}{(\rho+z)}} = \frac{M_y z}{K_1(\rho+z)}$$

where the value K_1 of the integral, determined in Appendix 4, is

$$K_1 = \int_\Omega \frac{z^2 d\omega}{r} = b\rho\left[\rho \log \frac{\rho+\frac{h}{2}}{\rho-\frac{h}{2}} - h\right]$$

Here again this normal stress will be a maximum in absolute value at the intrados fibre.

C. Bending around the z axis due to the Component Couple M_z. The stress component is as above given by the following equation:

$$X_x'' = \frac{M_z y}{r\int_\Omega \frac{y^2 d\omega}{r}} = \frac{M_z y}{K_2(\rho+z)}$$

where the integral K_2 (Appendix 4) has the following value:

$$K_2 = \int_\Omega \frac{y^2 d\omega}{r} = \frac{b^3}{12} \log \frac{\rho+\frac{h}{2}}{\rho-\frac{h}{2}}.$$

This stress component is maximum for $z=0$ and $y=\pm b$.

Limitation of the Stresses

1. *The Normal and Tangential Stress Components.*

The maximum value of the stress components may be limited separately. The maximum tangential stresses X_y and X_z have already been determined. The maximum of the normal stress X_x can be determined as follows:

At each of the points of any of the sections there exists a total normal stress X_x, which is equal to the sum of the stresses

$$X_x' + X_x'' = \frac{M_y z}{K_1(\rho+z)} + \frac{M_z y}{K_2(\rho+z)}$$

where the constants K_1 and K_2 have different values for the different cross sections.

Noting that for a constant z the second term of the above equation will increase as y increases, it is seen that for a given value of z, the point of maximum stress will lie on the periphery.

The total normal stress X_x at any point of the periphery is given by:

$$X_x = \frac{M_y z}{K_1(\rho+z)} + \frac{M_z y}{K_2(\rho+z)}$$

The ordinate z of the maximum of the normal stress X_x may be determined by equating to zero the first derivative of this expression with respect to z. The values of those maxima are obtained by introducing this value of z into the original equations.

Except when there exists a simple relation between the couples M_y and M_z, the above method leads to great difficulties.

The following method is therefore proposed:

We may explore the values of the normal stresses X_x throughout one quadrant of the section. If M_y and M_z have the same sign, the maximum negative stress component X_x' will occur at point d (Fig. 10) having as value:

$$X_x' = -\frac{M_y}{K}\frac{a}{\rho-a}$$

and the maximum negative stress component X_x'' will occur at point a. We thereby see that the point of maximum negative normal stress X_x lies on the periphery of any of the quadrants.

Likewise it can be readily shown that the maximum normal stress X_x of either sign will always be found in any one of the quadrants, independently of the sign of the moments M_y and M_z, and consequently the exploration of the values of this normal stress X_x throughout one of the quadrants will suffice.

Let such a quadrant be divided into a given number of small areas, their size and number depending on the degree of accuracy to be attained. For each of those regions, the stresses X_x' and X_x'' will be calculated and the values plotted as ordinates on a diagram where the values of z are in abscissa (Fig. 15). The

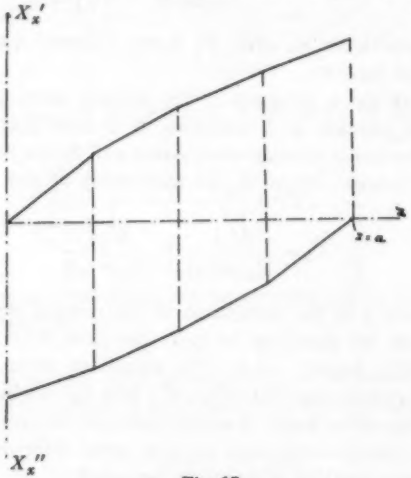


Fig. 15

points obtained that way may be joined by straight lines and the maximum of the normal stress X_x of both signs thereby obtained with any desired approximation.

2. *The Principal Stresses.*

If one desires to limit either the magnitude of the principal stresses and thereby the magnitude of the maximum normal stress (method of Lamé), or the greatest difference of the principal stresses (method of Coulomb and Tresca), the values of the principal stresses result from the following:

The general equation relating the principal stresses ϕ to the stress components here reduces to:

$$\phi^3 - X_x\phi^2 - (X_z^2 + X_y^2)\phi = 0$$

and the roots give the following values for the principal stresses:

$$\phi_1 = 0$$

$$\phi_2 = \tfrac{1}{2}X_x + \tfrac{1}{2}\sqrt{X_x^2 + 4(X_z^2 + X_y^2)}$$

$$\phi_3 = \tfrac{1}{2}X_x - \tfrac{1}{2}\sqrt{X_x^2 + 4(X_z^2 + X_y^2)}.$$

The directions of the principal stresses result from the equations of elastic equilibrium of the elementary right angle tetrahedron whose oblique face is a plane of principal stress having as direction cosines the values l, m, n. Those equations of elastic equilibrium here reduce to the following:

$$\begin{aligned} l(X_x - \phi) + m X_y + n X_z &= 0 \\ l X_y - m\phi &= 0 \\ l X_z - n\phi &= 0 \end{aligned}$$

with the condition equation

$$l^2 + m^2 + n^2 = 1.$$

The above equations yield the following values for the direction cosines of the three planes of principal stress:

$$l = \frac{\phi}{\sqrt{\phi^2 + X_y^2 + X_z^2}}$$

$$m = \frac{X_y}{\sqrt{\phi + X_y^2 + X_z^2}}$$

$$n = \frac{X_z}{\sqrt{\phi + X_y^2 + X_z^2}}$$

where ϕ is successively replaced by the values found for ϕ_1, ϕ_2, ϕ_3.

3. *The Maximum Deformation and the Corresponding Fictitious Stress.*

If one desires to limit the deformation e (method of Poncelet and de St. Venant), or the corresponding fictitious stress Ee (E being Young's modulus of elasticity), the maximum deformation may be obtained as follows:

When the three normal stress components X_x, Y_y, Z_z, and the three tangential stress components X_y, X_z, Y_z, have values different from zero, the deformation e in a direction l, m, n is given by:

$$e=l^2e_{xx}+m^2e_{yy}+n^2e_{zz}+mle_{xy}+nle_{xz}+mne_{yz}$$

where the six strain components are related to the six stress components by the stress-strain relations for isotropic bodies, namely:

$$e_{xx}=\frac{1}{E}\left[X_x-\sigma(Y_y+Z_z)\right]$$

$$e_{yy}=\frac{1}{E}\left[X_y-\sigma(X_x+Z_z)\right]$$

$$e_{zz}=\frac{1}{E}\left[Z_z-\sigma(X_x+Y_y)\right]$$

$$e_{yz}=\frac{1}{\mu}Y_z$$

$$e_{xy}=\frac{1}{\mu}X_y$$

$$e_{xz}=\frac{1}{\mu}X_z$$

where μ is the rigidity and equal to $\frac{E}{2(1+\sigma)}$. By substitution and simplification, the deformations e here become:

$$e=\frac{1}{E}\left\{\left[l^2(1+\sigma)-\sigma\right]X_x+2l(1+\sigma)(mX_y+nX_z)\right\}$$

The values of l, m, n, for which e becomes a maximum are determined by equating to zero the partial derivatives of the

expression for e with respect to l, m, n. The maximum deformation e is obtained by introducing those values in the original expression for e.

Conclusion

A solution for the torsional problem of curved beams is reached for the sections which have been considered, with no further limitations than those existing in the solutions of the corresponding torsional problem for straight members and the postulation of the principle of superposition. The method can be similarly applied wherever a solution of the problem of bending of the curved beam and the problem of torsion of the straight beam is available.

If the applied forces are not pure couples, they can in each section be reduced to a couple and a force. This force will in general cause in the further sections an additional bending around the y and z axes which is to be added to the values of M_y and M_z.

This method can also be used for non-prismatic beams provided the variations of profile are sufficiently slow to make the disturbance, caused thereby on the stress distribution, negligible.

APPENDIX 1

Calculation of $\int_\omega \frac{z^2 d\omega}{\rho+z}$ for circular section.

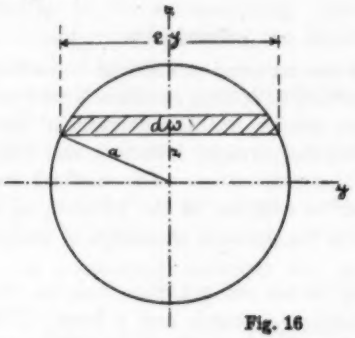


Fig. 16

$$\int_\omega \frac{z^2 d\omega}{\rho+z} = \int_{-a}^{+a} \frac{z^2 . 2y\, dz}{\rho+z} = 2\int_{-a}^{+a} \frac{z^2\sqrt{a^2-z^2}}{\rho+z}\, dz = K_1$$

where $d\omega = 2y\, dz$ and a the radius (Fig. 16). Introducing

$$z = a \cos t:$$

$$K_1 = 2a^4 \int_0^\pi \frac{\cos^2 t \sin^2 t}{\rho+a\cos t} dt = 2a^4 \int_0^\pi \frac{\cos^2 t}{\rho+a\cos t} + 2a^4 \int_0^\pi \frac{\cos^4 t}{\rho+a\cos t}\, dt.$$

On the other hand:

$$\frac{\cos^2 t}{\rho+a\cos t} = \frac{1}{a^2}\left[\frac{(a\cos t+\rho)(a\cos t-\rho)+\rho^2}{\rho+a\cos t}\right]$$

$$= \frac{\cos t}{a} - \frac{\rho}{a^2} + \frac{\rho^2}{a^2}\frac{1}{\rho+a\cos t}$$

$$\frac{\cos^4 t}{\rho+a\cos t} = \frac{\cos^3 t}{a} - \frac{\rho\cos^2 t}{a^2} + \frac{\rho^2}{a^2}\frac{\cos^2 t}{\rho+a\cos t}.$$

Hence:

$$K_1 = 2a^4 \int_0^\pi \frac{\cos^2 t}{\rho+a\cos t} dt + 2a^3 \int_0^\pi \cos^3 t\, dt - 2\rho a^2 \int_0^\pi \cos^2 t\, dt$$

$$+ 2\rho^2 a^2 \int_0^\pi \frac{\cos^2 t}{\rho+a\cos t} dt$$

or:

$$K_1 = 2a^2(a^2+\rho^2)\int_0^{\pi}\frac{\cos^2 t}{\rho+a\cos t}\,dt + 2a^3\int_0^{\pi}\cos^3 t\,dt - 2a^2\rho\int^{\pi}\cos^2 t\,dt$$

Calculating separately the different terms of the integral:

$$\int_0^{\pi}\frac{\cos^2 t}{\rho+a\cos t}\,dt = -\frac{\pi\rho}{a^2} + \frac{\rho^2}{a^2}\int_0^{\pi}\frac{dt}{\rho+a\cos t} = -\frac{\pi\rho}{a^2} + \frac{\rho^2}{a^2}\frac{\pi}{\sqrt{\rho^2-a^2}}$$

$$\int_0^{\pi}\cos^3 t\,dt = 2\int_0^{\pi}\sin^2 t\cos t\,dt = 2\int_0^{\pi}\cos t\,dt - 2\int_0^{\pi}\cos^3 t\,dt$$

$$= \frac{2}{3}\int_0^{\pi}\cos t\,dt = 0$$

Also:

$$\int_0^{\pi}\cos^2 t\,dt = \frac{1}{2}\int_0^{\pi}dt = \frac{\pi}{2}.$$

Hence:

$$K_1 = 2\pi\rho(\rho^2+a^2)\left(\frac{\rho}{\sqrt{\rho^2-a^2}} - 1\right) - \pi a^2\rho.$$

APPENDIX 2

Calculation of $\int_\Omega \frac{y^2 d\omega}{\rho+z}$ for circular section.

$$\int_\Omega \frac{y^2 d\omega}{\rho+z} = \int_{-a}^{+a}\frac{dz}{z+\rho}\int_{-\sqrt{a^2-z^2}}^{+\sqrt{a^2-z^2}} y^2 dy = K_2$$

As:

$$\int_{-\sqrt{a^2-z^2}}^{+\sqrt{a^2-z^2}} y^2 dy = \frac{2}{3}(a^2-z^2)^{3/2}.$$

And introducing $z = a\cos t$.

$$K_2 = \frac{2}{3}a^4\int_0^{\pi}\frac{\sin^4 t}{\rho+a\cos t}\,dt$$

$$= \frac{2}{3}a^4\int_0^{\pi}\frac{\sin^4 t}{\rho+a\cos t}\,dt - \frac{2}{3}a^4\int_0^{\pi}\frac{\sin^2 t\cos^2 t}{\rho+a\cos t}\,dt$$

Having:

$$\int_0^{\pi} \frac{\sin^2 t}{\rho + a\cos t}\,dt = \int_0^{\pi} \frac{dt}{\rho + a\cos t} - \int_0^{\pi} \frac{\cos^2 t\,dt}{\rho + a\cos t}$$

the different terms have been calculated in Appendix 1 and we consequently can write:

$$K_2 = \tfrac{2}{3}\pi a^2\left(\rho - \sqrt{\rho^2 - a^2}\right) - \tfrac{1}{3}\left[2\pi\rho\left(\rho^2 + a^2\right)\left(\frac{\rho}{\sqrt{\rho^2 - a^2}} - 1\right) - \pi a^2\rho\right]$$

APPENDIX 3

Elliptic cross section.

A.

$$K_1 = \int_0 \frac{z^2 d\omega}{\rho + z} = \frac{2b}{a}\int_{-a}^{+a} \frac{z^2\sqrt{a^2 - z^2}}{\rho + z}\,dz.$$

where a and b are the semi major and minor axes.

Introducing $z = a\cos t$

$$K_1 = \frac{b}{a}\,.\,2a^4\int_0^{\pi} \frac{\sin^2 t\cos^2 t}{\rho + a\cos t}\,dt.$$

As (Appendix 2):

$$2a^4\int_0^{\pi} \frac{\sin^2 t\cos^2 t}{\rho + a\cos t}\,dt = 2\pi\rho(\rho^2 + a^2)\left(\frac{\rho}{\sqrt{\rho^2 - a^2}} - 1\right) - \pi a^2\rho$$

$$K_1 = \frac{2\pi b\rho}{a}(a^2 + \rho^2)\left(\frac{\rho}{\sqrt{\rho^2 - a^2}} - 1\right) - \pi ab\rho.$$

B.

$$K_2 = \int_0 \frac{y^2 d\omega}{\rho + z} = \int_{-a}^{+a} \frac{dz}{\rho + z}\int_{-\frac{2b}{a}\sqrt{a^2 - z^2}}^{+\frac{2b}{a}\sqrt{a^2 - z^2}} y^2 dy.$$

As:

$$\int_{-\frac{2b}{a}\sqrt{a^2 - z^2}}^{+\frac{2b}{a}\sqrt{a^2 - z^2}} y^2 dy = \frac{16b^3}{3a^3}\left(a^2 - z^2\right)^{3/2}$$

$$K_2 = \frac{16b^3}{3a^3}\left\{\pi a^2(\rho - \sqrt{\rho^2 - a^2}) - (\rho^2 + a^2)\left(\frac{\pi\rho^2}{\sqrt{\rho^2 - a^2}} - \pi\rho\right) + \pi a^2\rho\right\}.$$

APPENDIX 4

Rectangular cross section.

A.
$$K_1=\int_\Omega \frac{z^2 d\omega}{\rho+z}=b\int_{-\frac{h}{2}}^{+\frac{h}{2}} \frac{z^2 dz}{\rho+z}$$

$$=b\int_{-\frac{h}{2}}^{+\frac{h}{2}} \left(z-\rho+\frac{\rho^2}{\rho+z}\right)dz.$$

Hence

$$K_1=b\rho\left[\rho \log \frac{\rho+\frac{h}{2}}{\rho-\frac{h}{2}}-h\right].$$

B.
$$K_2=\int_\Omega \frac{y^2 d\omega}{\rho+z}=\int_{-\frac{h}{2}}^{+\frac{h}{2}} \frac{dz}{\rho+z}\int_{-\frac{b}{2}}^{+\frac{b}{2}} y^2 dy.$$

As

$$\int_{-\frac{b}{2}}^{+\frac{b}{2}} y^2 dy=\frac{b^3}{12}$$

$$K_2=\frac{b^3}{12}\log\frac{\rho+\frac{h}{2}}{\rho-\frac{h}{2}}.$$

September, 1924.

A TAYLOR'S EXPANSION OF A DETERMINANT

By LEPINE HALL RICE

In the course of some work with determinants having binomial elements, the writer noticed the possibility of using the well-known development of such a determinant as a sum of determinants having monomial elements to express the successive terms of a Taylor's expansion of any determinant regarded as a function of all its elements. The literature presents several closely related results, from which this one seems an almost obvious deduction, but as the writer does not find it given, he ventures to publish it because, for one thing, it exhibits each term of the Taylor's expansion in a form suitable for study in respect to the part played by each individual element and by each row and column of the determinant. One may here remark that there is need of such types of expression in other parts of the theory of determinants where conciseness has been gained but elemental structure obscured.

Let

$$\Delta = \begin{vmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ x_{n1} & x_{n2} & \cdots & x_{nn} \end{vmatrix} = \begin{vmatrix} a_{11}+h_{11} & a_{12}+h_{12} & \cdots & a_{1n}+h_{1n} \\ a_{21}+h_{21} & a_{22}+h_{22} & \cdots & a_{2n}+h_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1}+h_{n1} & a_{n2}+h_{n2} & \cdots & a_{nn}+h_{nn} \end{vmatrix},$$

where all the elements have been represented as variables but where any of them, disposed in any manner, may be constants, x's becoming a's and the corresponding h's zeros. Arrange the monomial development as follows: Put first the determinant $A \equiv |a_{ij}|$; next a group of n determinants each formed from A by replacing one row of a's by the corresponding row of h's, this group being denoted by ΣA_1; next a group ΣA_2 of determinants, $\binom{n}{2}$ in number, formed by similarly replacing two rows

of a's by h's; and so on, until finally there comes a single determinant $A_n = |h_{ij}|$. Thus

$$\Delta = A + \Sigma A_1 + \Sigma A_2 + \ldots + A_n. \tag{1}$$

Beside this let us place a Taylor's expansion of

$$\Delta \equiv f(x_{11}, x_{12}, \ldots, x_{nn}),$$

viz.,

$$\begin{aligned}\Delta = {} & f(a_{11}, \ldots, a_{nn}) \\ & + \left(h_{11}\frac{\partial}{\partial x_{11}} + \ldots + h_{nn}\frac{\partial}{\partial x_{nn}}\right) f(a_{11}, \ldots, a_{nn}) \\ & + \frac{1}{2!}\left(h_{11}\frac{\partial}{\partial x_{11}} + \ldots + h_{nn}\frac{\partial}{\partial x_{nn}}\right)^2 f(a_{11}, \ldots, a_{nn}) \\ & + \ldots .\end{aligned} \tag{2}$$

Our result now is that *each term of* (1) *is equal to the corresponding term of* (2). For it scarcely need be pointed out that

$$\frac{\partial^r}{\partial x_{\alpha_1\beta_1}\, \partial x_{\alpha_2\beta_2} \ldots \partial x_{\alpha_r\beta_r}} f$$

vanishes unless all the α's are different and all the β's are different, and when they are different occurs $r!$ times, so that what survives of the $(r+1)$th term of (2) is

$$\frac{1}{r!} \sum_{\alpha} \sum_{\beta} r! \frac{\partial^r}{\partial x_{\alpha_1\beta_1} \ldots \partial x_{\alpha_r\beta_r}} f(a_{11}, \ldots, a_{nn}) = \Sigma A_r.$$

Inasmuch as determinants (of any signancy) and permanents of more than two dimensions are linear homogeneous functions of each layer (tranche) of any direction, a result like the foregoing holds for all determinants and permanents, the word "row" being replaced by the phrase "layer of any assigned direction."

SOMMERFELD'S THEORY OF FINE STRUCTURE FROM THE STANDPOINT OF GENERAL RELATIVITY[1]

By M. SANDOVAL VALLARTA

1. It is well known that the theory of the fine structure of the spectral lines of hydrogen, based essentially on the familiar formula for the variation of mass with velocity

$$m = m_0/\sqrt{1-(v/c)^2} \qquad (1)$$

where m is the mass at the uniform translation velocity v; m_0, the rest-mass and c the velocity of propagation of light in vacuum, is due to Sommerfeld[2], by whom it was proposed in 1916. Important applications of this theory to the general systematization of Röntgen spectra are also due to him. His fundamental conclusions were confirmed experimentally in the same year by Paschen[3], who found as a result of his observations on the spectrum of ionized helium He^+ a value of the fundamental hydrogen doublet ($0.3645\ cm^{-1}$) wholly in agreement with the theoretical value calculated from Sommerfeld's theory. More recently, however, Gehrcke and Lau[4] and other experimenters, operating directly on hydrogen at very low temperatures, have not confirmed Paschen's value, but obtained a lower figure for the magnitude of the above-mentioned doublet ($0.293\ cm^{-1}$) which, perhaps by a curious coincidence, verifies the theoretical value which would be obtained for this doublet if instead of using formula (1) for the variation of electronic mass with velocity, we were to use the formula given by Abraham[5] for the variation of the transversal mass of his rigid electron. From a theoretical standpoint, Lenard[6] has pointed out, correctly as we believe, that the experimental confirmation of Sommerfeld's theory does not imply, in contradiction

[1] Presented before the "Sociedad Cientifica Alzate," City of Mexico, August, 1924, and before the Pan-American Scientific Congress, Lima, December, 1924.

[2] A. Sommerfeld, Ann. d. Phys., Vol. 51, p. 53, 1916, or "Atombau und Spektrallinien," p. 570, Braunschweig, 1922.

[3] F. Paschen. Ann. d. Phys., Vol. 50, p. 901, 1916.

[4] E. Gehrcke & A. Lau. Ann. d. Phys., Vol. 67, p. 388, 1922.

[5] M. Abraham. "Theorie der Elektrizität," Vol. 2, p. 170; Leipzig, 1920. Cf. also Sommerfeld "Atombau und Spektrallinien," p. 644.

[6] Lenard's footnotes to J. Soldner's memoir in Ann. d. Phys., Vol. 65, p. 593, 1921.

to the latter's statement, any confirmation of the assumed fact that the laws derived from restricted relativity govern electronic motions in the atom, but rather a confirmation of any theory which leads to a law of variation of electronic mass with velocity identical to that expressed by formula (1). In fact, it is quite important to emphasize that not only the restricted theory leads to formula (1), but the latter can be deduced easily from considerations which have nothing in common with relativity, as is well known since Hasenöhrl and has been aptly brought out later by Lenard. It follows that it cannot be maintained that there exists a causal relationship between the restricted relativity theory and Sommerfeld's theory of fine structure.

More generally, the author of the present note has pointed out[7], that there can neither be any causal relationship between the metric of the space-time of the material field of a charged massic particle and the above-mentioned theory of Sommerfeld. Lenard has further pointed out that it is not easy to understand *a priori* how it is that the restricted theory of relativity is able of itself to account wholly for the motion of an electron around an atomic nucleus, whereas recourse must be had to general relativity to explain in detail the motion of a planet around the sun. It is well known that, in the case of Mercury for instance, the law of variation of mass with velocity expressed by (1) gives an advance of the perihelion of 7″ per century, whereas the Einstein law of gravitation gives 42.9″ and observation, according to the best published data to date, 41″.[8] This objection has already been discussed by Laue[9].

In the present paper we intend to show that, if we keep in mind the numerical relations between the magnitudes of the electronic mass and charge, on the one hand, and the nuclear mass and charge, on the other, it is possible to obtain the relativistic Keplerian conditioned-periodic orbits of Sommerfeld's theory starting from considerations of general relativity. In the

[7] Doctorate thesis "Bohr's Atomic Model from the Standpoint of the General Theory of Relativity and of the Calculus of Perturbations," Massachusetts Institute of Technology, May, 1924, p. 13.
[8] Grossmann. Zeits. für Phys., Vol. 5, p. 280, 1922.
[9] M. von Laue, Ann. d. Phys., Vol. 66, p. 284, 1921; or "Die Relativitätstheorie," Vol. 2, p. 238, Braunschweig, 1921.

following pages we shall see that if we keep in mind these relations as they exist in the hydrogen atom and in hydrogenic ions, of the type He^+, Li^{++}, etc., we come to the conclusion that in such cases the material field of the nucleus is almost Euclidean (Laue conditions) and almost static. As a preliminary, it might not be altogether out of the way to review briefly how the orbital equation is obtained in Sommerfeld's theory.

2. Consider an atomic nucleus of mass M and charge E, around which moves an electron of rest-mass m_0 and charge $-e$. Assume — an assumption which is true of the hydrogen atom and of every other atom and ion, without exception — that $m_0 << M$; we may then neglect the nuclear motion about the common center of gravity of the system nucleus-electron and it is easy to show that then we may likewise neglect the Newtonian force due to the masses m, M when compared with the Coulombian force due to the charges $-e$, E. Let us choose a polar coördinate system r, ϕ whose origin $r=0$ coincides with the nucleus. It is well known that in the dynamics of restricted relativity the momenta or impulses $p_1 \ldots p_n$ corresponding to the n coördinates $q_1 \ldots q_n$ which define a conservative holonomic dynamical system of n degrees of freedom are defined by means of a kinetic potential L,

$$L = m_0 c^2 \left(1 - \sqrt{1-\beta^2}\right) \tag{2}$$

($\beta = v/c$) as follows

$$p_1 = \frac{\partial L}{\partial \dot{q}_1} \ldots p_n = \frac{\partial L}{\partial \dot{q}_n} \tag{3}$$

while the kinetic energy T is defined by the expression

$$T = m_0 c^2 [(1-\beta^2)^{-\frac{1}{2}} - 1]. \tag{4}$$

It is important to observe that no relativistic consideration, outside of the law of variation of mass with velocity given by (1), enters into the derivation of expressions (2) and (4). $\beta = v/c$ is a function of the coördinates $q_1 \ldots q_n$ and the velocities $\dot{q}_1 \ldots \dot{q}_n$.

The first fundamental form in the plane in polar coördinates is, as known,

$$ds^2 = dr^2 + r^2 d\phi^2$$

hence

$$\beta^2=\frac{1}{c^2}\left(\frac{ds}{dt}\right)^2=\frac{1}{c^2}\left(\dot{r}^2+r^2\dot{\phi}^2\right) \quad . \quad . \quad . \tag{5}$$

and

$$L=m_0c^2\left[1-\left(1-\frac{\dot{r}^2}{c^2}-r^2\frac{\dot{\phi}^2}{c^2}\right)^{\frac{1}{2}}\right]$$

and, by (3),

$$\left.\begin{aligned} p_r&=\frac{\partial L}{\partial \dot{r}}=m_0(1-\beta^2)^{-\frac{1}{2}}\,\dot{r}=m\dot{r}\\ p_\phi&=\frac{\partial L}{\partial \dot{\phi}}=m_0(1-\beta^2)^{-\frac{1}{2}}r^2\dot{\phi}=mr^2\dot{\phi}\end{aligned}\right\} \tag{6}$$

which, except for the meaning of m (mass at the velocity v), are formally identical to the corresponding quantities in classical dynamics.

We now proceed to find the Hamiltonian function $H(p, q)$ of the system in question. Since the latter is conservative by hypothesis, according to Bohr's first postulate, while it stays in a quantized orbit, we may write

$$T+V=W=\text{const.} \tag{7}$$

V being the potential energy and W the energy-constant of the system. We have, by (5) and (6):

$$\beta^2=\frac{1}{m^2c^2}\left(p_r^2+\frac{1}{r^2}p_\phi^2\right)$$

whence

$$(1-\beta^2)^{-\frac{1}{2}}=\left[1+\frac{1}{m_0^2c^2}\left(p_r^2+\frac{1}{r^2}p_\phi^2\right)\right]^{\frac{1}{2}}.$$

Substituting this last expression in (4), introducing in the latter the energy-integral (7) and remembering that $V=-eE/r$, we get at once the required Hamiltonian function:

$$H(p, q)=1+\frac{1}{m_0^2c^2}\left(p_r^2+\frac{1}{r^2}p_\phi^2\right)-\left(1+\frac{W+\frac{eE}{r}}{m_0c^2}\right)^2=0 \tag{8}$$

which may also be written, after a slight algebraic transformation

$$p_r^2+\frac{1}{r^2}p_\phi^2=2m_0W+\frac{2m_0eE}{r}+\frac{1}{c^2}\left(W+\frac{eE}{r}\right)^2. \qquad (9)$$

From this we might evidently obtain the equation of the orbit through Hamilton-Jacobi's equation and Jacobi's celebrated theorem, but it is more convenient and incomparably easier to proceed as follows:

We have, obviously, from (6):

$$\frac{p_r}{p_\phi}=\frac{1}{r^2}\frac{dr}{d\phi}=-\frac{du}{d\phi}$$

where $u=1/r$. It follows that $p_r=-p_\phi\dfrac{du}{d\phi}$ and substituting in (9) we get at once the required equation of the electronic orbit:

$$\left(\frac{du}{d\phi}\right)^2=\frac{W}{p_\phi^2}\left(2m_0+\frac{W}{c^2}\right)+\frac{2eE}{p_\phi^2}\left(m_0+\frac{W}{c^2}\right)u+\left(\frac{e^2E^2}{c^2p_\phi^2}-p_\phi^2\right)u^2 \qquad (10)$$

where it is to be noted that from (8),

$$\dot{p}_\phi=\frac{\partial H(p,q)}{\partial\phi}=0 \qquad \text{whence } p_\phi=\text{const.}$$

As the right-hand member $f(u)$ of (10) is a rational polynomial of the second degree in u, the integral of the orbital equation

$$\phi=\int\frac{du}{\sqrt{f(u)}}$$

can be expressed in circular functions of u, as is well known. Otherwise, differentiating (10) with respect to ϕ we obtain the equation of the orbit in its more familiar form:

$$\frac{d^2u}{d\phi^2}=\frac{eE}{p_\phi^2}\left(m_0+\frac{W}{c^2}\right)+\left(\frac{e^2E^2}{c^2p_\phi^2}-1\right)u \qquad (11)$$

whose integral is, in fact,

$$u=A\cos\alpha\phi+B\sin\alpha\phi+C$$

where

$$\alpha^2=1-\frac{e^2E^2}{c^2p_\phi^2}, \qquad C=\frac{m_0eE}{\alpha^2p_\phi^2}\left(1+\frac{W}{m_0c^2}\right)$$

and A, B are integration constants. The presence of the factor

a in the general integral causes the relativistic advance of the perihelion, resulting in the fact that the relativisitic Keplerian orbit is not closed but conditioned-periodic, as shown in 1916 by Sommerfeld and by Epstein.

3. As stated in the introduction, we now intend to prove that if we take into account the numerical relations between charges and masses, such as they exist in the hydrogen atom and in hydrogenic ions, equation (10) defining the electronic orbit, which was obtained in the previous paragraph from considerations of restricted relativity which are not *a priori* applicable to the case on hand, can also be obtained from general considerations based on the general theory of relativity. This result was first established, although not quite completely, by Jaffé[10] and studied again by the author of the present paper.

As a preliminary we have to solve the following problem: to determine the metric fundamental form of the material field of the nucleus. The problem has been already dealt with from a more or less special standpoint by several writers, among whom we might mention Nordström[11], Longo[12], Jeffery[13], and Weyl[14], and from the more general standpoint of the generalized Hamilton Principle by De Donder and Vanderlinden[15]. In view of the fundamental importance of this problem, we might, perhaps properly, add a few words on the subject.

We consider the case of an isolated charged massic sphere, at rest, and we study its field such as it appears "*im kleinem*" to an infinitesimal observer. Under these conditions, this sphere gives rise to a static sphero-symmetrical material field and the same applies to the distribution of charge over this sphere. It follows that all the components of the four-vector electromagnetic potential ϕ_k are identically zero except the component ϕ_4 which depends only on the radial variable introduced by Schwarzschild and

[10] G. Jaffé. Ann. d. Phys., Vol. 67, p. 215, 1922.
[11] G. Nordström. Verslag Koninklijke Akademie van Wetenschappen te Amsterdam., Vol. 26, p. 1238, 1918.
[12] C. Longo. Il Nuovo Cimento., Vol. 15, p. 191, 1918.
[13] G. B. Jeffery. Proc. of the Royal Soc. of London, Vol. 99, p. 123, 1921.
[14] H. Weyl. "Raum, Zeit, Materie," p. 257, Berlin 1923; hereafter referred to as l.c.
[15] De Donder, "La gravifique einsteinienne," pp. 22 and 151, Paris, 1921

may therefore be physically likened to the ordinary electrostatic potential. The essential difficulty of the problem on hand lies in the fact that the massic and electric fields influence each other and cannot be determined independently.

Since the material field under consideration is static and has spherical symmetry, it follows by definition that the metric of its space-time is defined in turn by the quadratic differential form

$$ds^2 = \alpha(r)dr^2 + \psi(r)[d\theta^2 + \sin^2\theta d\phi^2] + \gamma(r)dt^2 \qquad (12)$$

which transformed to Schwarzschild variables, by the transformation:

$$x_1 = \tfrac{1}{3}r^3$$
$$x_2 = -\cos\theta$$
$$x_3 = \phi$$
$$x_4 = t$$

may be written,

$$ds^2 = -f_1 dx_1^2 - f_2 \frac{dx_2^2}{1-x_2^2} - f_3(1-x_2^2)dx_3^2 + f_4 dx_4^2 \qquad (13)$$

where the f's are functions of x_1 only. In what follows we shall attempt to give a simplified treatment of the methods used by De Donder to find the expression. for the f's which corresponds to the external field of the material sphere already defined above.

We start from the generalized Hamilton principle which, in the form suggested by De Donder, consists in *assuming* that, for every physical process and at every region of the four-dimensional space-time manifold, we have,

$$\delta \int_R (a + bK + \Lambda)\sqrt{-g}\, dx_1 dx_2 dx_3 dx_4 = 0 \qquad (14)$$

where a and b are universal constants, K the total Gaussian curvature of the space-time defined by $ds^2 = g_{\mu\nu}dx_\mu dx_\nu$[16], Λ is the energy-density, a function characterizing the space-time studied, which depends on the potentials $g_{\mu\nu}$, on their derivatives, and

[16] If $R_{\mu\nu}$ is the tensor obtained by contraction from the Riemann-Christoffel tensor, and $g^{\mu\nu}$ the fundamental contravariant tensor corresponding to $g_{\mu\nu}$, $K = g^{\mu\nu} R_{\mu\nu}$, as is well known.

also on the variables defining the electromagnetic field. The variation of (14) is to be taken only with respect to $g^{\mu\nu}$ and their derivatives inside the region R of space-time which is considered; on the boundary of this region the variation vanishes. g is as usual the determinant of the metric fundamental form and $\sqrt{-g}\, dx_1 dx_2 dx_3 dx_4$ is the element of volume in space-time. By taking the variation of each one of the terms in (14) we get, after laborious calculations of course, the general equations of the field[17]. Thus, the Hamilton principle as introduced by De Donder appears as a generalization of this principle such as used by Lorentz[18], Hilbert,[19], Weyl[20] and Einstein himself[21].

In the case considered by Einstein, that the universal constant a is zero, the general field equations are written[22]

$$2bR_{\mu\nu}\sqrt{-g} = T_{\mu\nu} = \tfrac{1}{2}g_{\mu\nu}T \tag{15}$$

where $T_{\mu\nu}$ is the covariant symmetric tensor of the electromagnetic and massic fields, and T the corresponding invariant. In the material field, the symmetrical tensor $T_{\mu\nu}$ is the sum of the tensors belonging to the massic and the electromagnetic fields: we shall call them $T^m_{\mu\nu}$ and $T^e_{\mu\nu}$ respectively.

As De Donder has shown, the ollowing relation holds between the tensor $T^e_{\mu\nu}$ and the derived electromagnetic tensors $E_{\mu a}$.

$$T^e_{\mu\nu} = -\sqrt{-g}\, g^{ab}(E_{\mu a}E_{\nu b} + E_{\mu a}{}'E_{\nu b}') \tag{16}$$

also $T^e = 0$.

If

$$E_{\mu a}{}' = \frac{\partial \phi_\mu}{\partial x_a} - \frac{\partial \phi_a}{\partial x_\mu} \tag{17}$$

is the ordinary electromagnetic tensor and ϕ_k the electromagnetic potential already mentioned, then the derived tensors $E_{\mu a}$ are

[17] T. De Donder, " La gravifique einsteinienne," p. 22 and p. 151, Paris, 1921. Hereafter referred to as l.c.

[18] H. A. Lorentz. Verslag Koninklijke Akademie van Wetenschappen te Amsterdam, Vol. 23, p. 1073; Vol. 24, p. 1389; Vol. 25, p. 468.

[19] D. Hilbert, Nachrichten der Gesellschaft der Wissenschaften zu Göttingen, p. 395, 1915; p. 53, 1917.

[20] H. Weyl. l.c. p. 229. Also Ann. d. Phys., Vol. 54, p. 117, 1917.

[21] A. Einstein. Sitzungsberichte der preussischen Akademie der Wissenschaften, Vol. 42, p. 1111, 1916.

[22] T. De Donder. l.c., p. 24.

the dualistics of the tensors $E_{\mu\alpha}'$ and can be derived from the latter by the formula

$$E_{\mu\nu} = -\frac{1}{\sqrt{-g}} \Sigma(-1)^{i+j} \begin{vmatrix} g_{\alpha\mu} & g_{\beta\mu} \\ g_{\alpha\nu} & g_{\beta\nu} \end{vmatrix} E_{ij}' \tag{18}$$

or conversely by using Buhl's symbolic determinants. In (18) the summation is extended to the six combinations of 1, 2, 3, 4 taken in pairs and α, β are the digits which remain from the former if in such digits we eliminate the two digits which enter into the combination we consider i, j $(i<j)$. We note at once that the electromagnetic symmetrical tensor $T_{\mu\nu}^{e}$ is a function of the ten gravitational potentials $g_{\mu\nu}$ and of the four electromagnetic potentials, ϕ_k.

The differential equations of the material field are given by (15). The value of the contracted Riemann-Christoffel tensor in a static sphero-symmetrical field is easily found. The only components which do not vanish are known to be

$$\left.\begin{aligned} R_{11} &= -\tfrac{1}{2}\frac{\partial}{\partial x_1}\left(f_1^{-1}\frac{\partial f_1}{\partial x_1}\right) + \tfrac{1}{4}f_1^{-2}\left(\frac{\partial f_1}{\partial x_1}\right)^2 + \tfrac{1}{2}f_2^{-2}\left(\frac{\partial f_2}{\partial x_1}\right)^2 + \tfrac{1}{4}f_4^{-2}\left(\frac{\partial f_4}{\partial x_1}\right)^2 \\ R_{22} &= R_{33} = \tfrac{1}{2}\frac{\partial}{\partial x_1}\left(f_1^{-1}\frac{\partial f_2}{\partial x_1}\right) - \tfrac{1}{2}\left(f_1 f_2\right)^{-1}\left(\frac{\partial f_2}{\partial x_1}\right)^2 - 1 \\ R_{44} &= -\tfrac{1}{2}\frac{\partial}{\partial x_1}\left(f_1^{-1}\frac{\partial f_4}{\partial x_1}\right) + \tfrac{1}{2}\left(f_1 f_4\right)^{-1}\frac{\partial f_4}{\partial x_1} \end{aligned}\right\} \tag{19}$$

The second members of (15) involve a massic tensor $T_{\mu\nu}^{m}$ and an electromagnetic tensor $T_{\mu\nu}^{e}$. The part involving the massic field offers no new difficulty and has already been treated in Schwarzschild's celebrated memoir.[23] We consider now the part involving the electric field. As already stated before, the only component of the tensor $E_{\mu\nu}'$ which does not identically vanish is

$$E_{14}' = -\frac{\partial \phi_4}{\partial x_1} \tag{20}$$

and substituting in (18) we find,

$$E_{23} = -\frac{1}{c} f_2^2 \frac{\partial \phi_4}{\partial x_1}$$

[23] K. Schwarzschild. Berichte der preussischen Akademie der Wissenschaften, p. 189, 1916; p. 424, 1916.

The remaining components vanish. Finally, substituting in (16) we get the symmetric electromagnetic tensor $T^e_{\mu\nu}$:

$$\left.\begin{aligned} T^e_{11} &= -\frac{1}{2c} f_1 f_2^2 \left(\frac{\partial \phi_4}{\partial x_1}\right)^2 \\ T^e_{22} &= T^e_{33} = \frac{1}{2c} f_2^3 \left(\frac{\partial \phi_4}{\partial x_1}\right)^2 \\ T^e_{44} &= \frac{1}{2c} f_2^2 f_4 \left(\frac{\partial \phi_4}{\partial x_1}\right)^2 \end{aligned}\right\} \tag{21}$$

The remaining components also vanish. Since the field has spherical symmetry we may set without loss of generality $\theta = \pi/2$ or what is the same, $x_2 = 0$ ($x_2 = -\cos\theta$) for greater simplicity.

These preliminaries are sufficient to enable us to solve the external problem, that is, for the region $r \geq a$ if a is the Schwarzschildian radius of the sphere with which alone we are concerned for the present. Remembering in fact that for this external field we have

$$\left.\frac{\partial \phi}{\partial x_1}\right|_{r \geq a} = -\frac{cE}{f_2^2} \qquad (E = \text{nuclear charge})$$

we obtain explicitly, substituting in the field equations (15) the values of the massic tensors,[24] the differential equations of the material field external to the nucleus:

$$\left.\begin{aligned} 2bcR_{11} &= -\tfrac{1}{2} f_1 f_2^{-2} E^2 \\ 2bcR_{22} &= 2bcR_{33} = \tfrac{1}{2} f_2^{-1} E^2 \\ 2bcR_{44} &= \tfrac{1}{2} f_1 f_2^{-2} E^2 \end{aligned}\right\} \tag{22}$$

It remains now to integrate the above system of partial differential equations. This integration has been effected by Vanderlinden[25] using a method similar to that given by Schwarzschild in his above-mentioned memoir.

[24] It is to be noted that the cosmological term of the second form of Einstein's law of gravitation has nothing to do with the case under consideration, because we are dealing only with "im kleinem" properties of the universe, such as they appear to an infinitesimal observer. In any case, the cosmological term can give rise only to small correction terms of the order of magnitude of that term. Cf. Eddington, Mathematical Theory of Relativity, p. 100.

[25] H. Vanderlinden. Bulletin de l'Académie Royale de Belgique, May, 1921.

Letting $R=(3x_1+\delta)^{1/3}$ where δ is an integration constant, the solution is written:

$$f_1=\left[R^4\left(1-\frac{2m}{R}+\frac{\epsilon^2}{R^2}\right)\right]^{-1}$$
$$f_2=R^2 \qquad (23)$$
$$f_4=c^2\left(1-\frac{2m}{R}+\frac{\epsilon^2}{R^2}\right)$$

where $2m$ is another integration constant to which, as in Schwarzschild's solution of the field outside the massic sphere, we ascribe the physical meaning of gravitational mass, and $\epsilon^2=-E^2/4b$ describes the contribution to the curvature of space-time of the gravitational effect of the electric charge.

Transforming back to the polar coordinates which were used to write the form (12), we have the metric fundamental form defining the space-time outside a material sphere:

$$ds^2=-\left(1-\frac{2m}{r}+\frac{\epsilon^2}{r^2}\right)^{-1}dr^2-r^2(d\theta^2+\sin^2\theta d\phi^2)+$$
$$+c^2\left(1-\frac{2m}{r}+\frac{\epsilon^2}{r^2}\right)dt^2 \qquad (24)$$

Obviously, if $\epsilon=0$ we fall back on Schwarzschild's external field which is a particular case of the problem of the material field.

The absence of empirical or semi-empirical data in the above method of solution must be very carefully emphasized. In Weyl's solution, for instance, the square of the electric field intensity is taken as the action-function of the electric field. To quote him textually:[36] "Als Wirkungsfunktion des elektrischen Feldes ist gemäss der klassischen Theorie das Quadrat des Feldbetrages zugrunde gelegt."

4. We now have to consider the following problem: What is the world-line of a "small" particle of charge $-e$ and mass m (electron) in the field of a "large" particle of charge E and mass M (nucleus)? We assume for the present that the mass m and the charge $-e$ are so small compared to M and E that the material

[36] H. Weyl, l.c., p. 257.

field of the nucleus is not perceptibly affected by the presence of the electron there. The problem has already been attacked from a more of less empirical standpoint by Jeffery,[27] discussed by Eddington[28] and studied from more general points of view by Weyl[29] and De Donder[30].

Why does an electron move when placed in an electromagnetic field? From an elementary standpoint, the answer is: it moves because of the existence of the ponderomotive force of the field. But, as Eddington has observed, this is nothing more than a simple verbal confusion: in reality electric " force " has nothing in common with mechanical " force " and describes, from the relativistic standpoint, a world condition totally different from that described by a " force " in the mechanical sense of the word. From what we might call a relativistic-empirical standpoint, the problem is formulated as follows: Find the tensor equation corresponding to the elementary equation of electrostatics:

$$m\frac{d^2r}{dt^2} = eE$$

where r and E are ordinary three-dimensional vectors and it is of course assumed that the field is static. Jeffery has suggested that such a tensor equation is,

$$m\left[\frac{du^\mu}{ds} + \left\{\begin{matrix}\alpha & \beta \\ & \mu\end{matrix}\right\} u^\alpha u^\beta\right] = E_\mu{}^\nu \, e \, u^\nu \qquad (25)$$

where $u^\mu = (dx)^\mu/ds$ is the generalized velocity (Eddington) or, perhaps better still, " the world-direction μ " (Weyl).

It is seen at once that the world-line of the charged particle is not a geodesic in space-time, owing to the existence of a second member in (27), that is, owing to the presence of the electric field, differing thus *fundamentally* in its behavior from that of the massic particle in a pure gravitational field. We now inquire with Eddington, what is the exploring particle trying to reveal to us about the physical structure of the universe by deviating from the

[27] G. B. Jeffery. Proc. of the Royal Society of London. Vol. 99, p. 123, 1921.

[28] A. S. Eddington, The Mathematical Theory of Relativity, p. 186, Cambridge, 1923. Hereafter referred to as l.c.

[29] H. Weyl., l.c., pp. 277–287.

[30] T. De Donder, l.c., p. 80.

geodesic path? Eddington's own profound examination of this problem leads to decidedly pessimistic results and shows just this: that if the existence of the electron contradicts our ordinary electromagnetic equations, in such a manner that an isolated electron at rest is, as far as we are able to study it at present, nothing short of a miracle, the same electron in an external field of force and endowed with the acceleration expressed by (25) is precisely the same miracle. To use Eddington's somewhat ironic expression " that is as far as the explanation goes." Weyl, on the other hand, is satisfied with proving that Einstein's field laws demand that the world-direction of a particle whose charge is e and whose mass is m vary as required by the equation

$$m\left(\frac{du_i}{ds}-\tfrac{1}{2}\frac{\partial g_{\alpha\beta}}{\partial x_i}u^\alpha u^\beta\right)=\frac{\kappa e}{2}E_{ki}u^k \qquad (25')$$

where $u_i=u^k g_{ik}$ and κ is a constant depending on the units used to express e. But it is clear that if the particle in question is an electron this does not advance Eddington's difficulty. Van den Berg's memoir on the relativistic theory of Poincaré's electron would certainly be a notable advance in the theory of the electron if it were not for the introduction of the so-called applied Poincaré tensor which is nothing but a substitute name for the non-Maxwellian system used by Eddington to explain the existence of the electron.

The present writer does not know for certain yet whether a purely geometric and general derivation of equation (25), based either on Weyl's theory of the electromagnetic field or on Eddington's generalization, is possible. *A priori* it seems that it ought to be so, for if the mechanical properties of a particle depend upon the non-integrability of direction in Riemann's continuum and are described by the ten gravitational potentials $g_{\mu\nu}$, while its electromagnetic characteristics depend on the non-integrability of length which are in turn determined by four functions ϕ_k in Weyl's non-Riemannian geometry, it must follow that this geometrical characteristic of the electromagnetic field appears physically as a difference between the behavior of the charged and the uncharged particle. In the same way that, for example, the non-integrability of direction appears physically to us in the

relativistic conditioned-periodic Keplerian orbit which is a geodesic of the massic field or in the geodesic precession discovered by Fokker and Schouten, so the non-integrability of length may cause the deviation between the geodesic orbit of the massic particle and the non-geodesic orbit of the charged particle. The analytical development of this idea is left for another occasion.

Using equation (25) or (25′) and the metric fundamental form (24), we proceed to the determination of the electronic orbit in the material field of the nucleus. That the nuclear field is not appreciably altered by the presence of the electron will be proved later. We have seen already that in a static sphero-symmetrical field the only components of E_{ik} which do not vanish are

$$E_{14}=-E_{41}=\frac{\partial\phi_4}{\partial x_1}\,.$$

The equation corresponding to the variable θ is independent of the electric field because the corresponding component of the electromagnetic tensor vanishes. Following Eddington's method, one finds

$$\frac{d^2\theta}{ds^2}+\frac{2}{r}\frac{dr}{ds}\frac{d\theta}{ds}-\sin\theta\cos\theta\left(\frac{d\phi}{ds}\right)^2=0. \tag{26}$$

An identical solution of this equation is $\theta=\pi/2$. It follows that if $\theta=\pi/2$ at any time, θ is always $\pi/2$ and consequently the electronic orbit is plane and the existence of the orbital plane is established. On this plane $d\theta=0$ and hence the linear element reduces to

$$ds^2=-\left(1-\frac{2m}{r}+\frac{\epsilon^2}{r^2}\right)^{-1}dr^2-r^2d\phi^2+c^2\left(1-\frac{2m}{r}+\frac{\epsilon^2}{r^2}\right)dt^2. \tag{27}$$

The equations corresponding to the variables t and ϕ depend on the electric field through the tensor E_{ik} and are respectively,

$$\frac{d^2t}{ds^2}+\frac{dr}{ds}\frac{dt}{ds}\frac{\partial}{\partial r}\log\gamma=\frac{eE}{mcr^2\gamma}\frac{dr}{ds} \tag{28}$$

$$mc\frac{d^2\phi}{ds^2}+\frac{2mc}{r}\frac{dr}{ds}\frac{d\phi}{ds}=0 \tag{29}$$

where

$$\gamma=1-\frac{2m}{r}+\frac{\epsilon^2}{r^2}\,.$$

The integration of (28) and (29) offers no difficulty and gives the first integrals which correspond to the energy-integral and the integral of angular momentum of classical dynamics:

$$\gamma\frac{dt}{ds}=C+\frac{eE}{mcr}=W+mc^2+\frac{eE}{mcr} \tag{30}$$

$$mcr^2\frac{d\phi}{ds}=p \tag{31}$$

C and p are here integration constants; we place C equal to the well-known expression of the total energy in relativistic dynamics, which will be justified in the next paragraph. From (27) we get at once,

$$c^2\gamma\left(\frac{dt}{ds}\right)^2-\gamma^{-1}\left(\frac{dr}{ds}\right)^2-r^2\left(\frac{d\phi}{ds}\right)^2-1=0 \tag{32}$$

and eliminating from this dt and ds by means of the integrals of energy and angular momentum (30) and (31), we get at once the orbital equation,

$$\left(\frac{du}{d\phi}\right)^2-\frac{W}{p^2c^2}-\frac{2mW}{p^2}+\frac{2meE}{p^2}\left(1+\frac{W}{mc^2}+\frac{\kappa c^2mM}{eE}\right)u-$$
$$-\left(1+\kappa\frac{m^2E^2}{p^2}-\frac{e^2E^2}{c^2p^2}\right)u^2+2\kappa Mu^3-\frac{\kappa E^2}{c^2}u^4=0 \tag{33}$$

where, as usually in dynamics, we have introduced the variable $u=1/r$ in place of r and used the constant κ to reduce everything to C. G. S. units; $\kappa=7.42\times10^{-29}$, *i.e.*, the gravitational constant divided by the square of the velocity of light.

The integral of (33) is, in terms of a single quadrature,

$$\phi=\int\frac{du}{\sqrt{\phi(u)}}$$

and as $\phi(u)$ is an entire polynomial of the fourth degree in u, the integration can be carried out in terms of elliptical functions of Jacobi or Weierstrass.[31] It is to be noted that the orbit of a massic particle in a pure gravitational field is a particular case of the above equation (33) and is obtained if in the latter we set $e=E=0$. A brief study of the orbits in the material field has

[31] See for ex. Whittaker and Watson's "Modern Analysis" par. 22-7. Cambridge, 1916.

been published by Weyl, but from the point of view of the applications to atomic theory, a complete study of these orbits, such as that given by De Jans for the case of the pure gravitational field is of relatively small importance, for the reasons given in the next paragraph.

Finally, differentiating (33) with respect to ϕ we obtain the equation of the orbit in the more convenient form:

$$\frac{d^2u}{d\phi^2}=\frac{meE}{p^2}\left(1+\frac{W}{mc^2}+\frac{\kappa c^2 mM}{eE}\right)-\left(1+\kappa\frac{e^2E^2}{p^2}-\frac{e^2E^2}{c^2p^2}\right)u+$$

$$+3\kappa Mu^2-\frac{2\kappa E^2}{c^2}u^3. \qquad (34)$$

5. We have now to consider the following two questions:

(*a*) In the hydrogen atom or in hydrogenic ions of the type He^+, Li^{++}, Be^{+++}, etc.; how does the metric of space-time in the material field of the nucleus differ from the euclidean metric in a Minkowski field? We have therefore to show that the two quantities

$$\alpha_1=\frac{2\kappa M}{R} \qquad \beta_1=\frac{\kappa E^2}{R^2} \qquad (35)$$

are both very small compared to 1. These are what we have called the Laue conditions[32] and measure the discrepancy between the actual curvature in the material field of the nucleus and the null field curvature of the Minkowskian field assumed in special relativity. Further, if the numerators of the quantities (35) are small enough, the Schwarzschildian radius r differs but little from the euclidean radius and we may take the latter for the former.

(*b*) In the above-mentioned cases, how does the *actual* material field of the nucleus differ from the *static* field assumed in the fundamental form (26)? A measure of this discrepancy is the ratio of the quantities α_1, β_1, (35) which are a measure of the curvature of the space-time outside the nucleus, to the quotient

$$\left(\frac{d\sigma}{dt}\right)^2 \div c^2\gamma$$

which measures the way in which the point-singularity due to the electron evolves in the pure nuclear field. Here $d\sigma^2$ is the

[32] M. von Laue. Ann. d. Phys. Vol. 66, p. 284, 1921.

space-interval (that is, involving the variables r, θ, ϕ only) which enters in the form (26), t is the time measured by a Minkowski observer. If the *actual* field is almost static and almost sphero-symmetrical, we must have

$$\left.\begin{aligned} \alpha_2 &= \frac{\alpha_1 c^2 \gamma}{\left(\frac{d\sigma}{dt}\right)^2} << 1 \\ \beta_2 &= \frac{\beta_1 c^2 \gamma}{\left(\frac{d\sigma}{dt}\right)^2} << 1 \end{aligned}\right\} \tag{36}$$

These conditions were first proposed in a somewhat different sense by Jaffé and first used in this sense by the writer[33]. They measure the discrepancy between the actual relativistic two-body problem which we have in reality and the one-body problem which has been assumed throughout this paper.

To prove that the conditions (35) and (36) are verified in the hydrogen atom it is only necessary to consider well known and generally accepted numerical relations. Obviously, the nearer the quantized orbit is to the nucleus, the greater weight is to be attached to these conditions. In Sommerfeld's nomenclature, the quantized orbit which is nearest the nucleus is characterized by the azimuthal quantum number $n_\phi = 1$ and the radial quantum number $n_r = 0$, *i.e.*, the first Bohr circle. We have for this orbit[34]

$$\begin{aligned} e &= 4.774 \times 10^{-10} \text{ statcoulombs} \\ E &= 4.774 \times 10^{-10} \text{ statcoulombs} \\ m &= 8.99 \times 10^{-28} \text{ grams} \\ M &= 1.649 \times 10^{-24} \text{ grams} \\ \left(\frac{d\sigma}{dt}\right)^2 \Big/ c^2 &= 7.29 \times 10^{-3} \text{ cm. (Sommerfeld's constant)} \\ r &= 0.532 \times 10^{-8} \text{ cm.} \end{aligned}$$

whence, by (35) and (36)

$$\alpha_1 = 4.61 \times 10^{-16}, \qquad \beta_1 = 6.62 \times 10^{-53}. \tag{37}$$

[33] Doctorate Thesis "Bohr's Atomic Model from the Standpoint of the General Theory of Relativity and of the Calculus of Perturbations," p. 32.
[34] A. Sommerfeld "Atombau und Spektrallinien," p. 9 and p. 262, Braunschweig, 1922.

It follows that Laue's conditions are verified to a very high degree of approximation in the field of the hydrogen nucleus. Consequently, as explained above, $\gamma = 1$ nearly and

$$\alpha_2 = 0.634 \times 10^{-49}, \quad \beta_2 = 0.923 \times 10^{-56} \tag{38}$$

and to all purposes the hydrogen atom may be treated as a one-body problem; further the material field of the nucleus is very nearly static and sphero-symmetrical. We also note in passing that it is quite legitimate to use the euclidean radius vector instead of the Schwarzschildian radius.

It is easy to see that these conditions are still more approximately satisfied for hydrogenic ions of the type He^+, Li^{++}, Be^{+++}, etc., but it must be very carefully emphasized that they are *not* applicable to atoms and ions of more than one electron owing to the interpenetration of the different orbits suggested by Bohr and his school. Under such conditions, the nuclear field is neither static nor sphero-symmetrical and from this fact can be derived a simple relativistic explanation of the existence of unmechanical orbits in the atom, which have been repeatedly pointed out by Bohr, Born and their collaborators, as has already been suggested by the present writer[35] and will be more fully presented shortly.

Concerning electronic orbits of *any* quantum numbers, Sommerfeld's analysis shows that the relativistic Keplerian orbit differs most from the classical Keplerian orbit for those orbits for which $n_\phi = 1$ and n_r is a large number. But even for the largest values of the radial quantum number which may occur in nature, our fundamental quantities α_1, β_1, α_2, β_2, stay of the same order of magnitude, as is easily verified.

From the above conclusions follows that all the terms depending on the curvature of space-time may be dropped from equation (36), that is, all the terms involving the factor κ. If we do so we obtain:

$$\frac{d^2u}{d\phi^2} = \frac{meE}{p^3}\left(1 + \frac{W}{mc^2}\right) - \left(1 - \frac{e^2E^2}{c^2p^2}\right)u \tag{11'}$$

which is precisely Sommerfeld's equation (11). It follows further

[35] M. Sandoval Vallarta, Journal of Mathematics and Physics, Vol. 3, p. 174, 1924.

that his theory of fine structure is identical to that derived from the general theory of relativity if we neglect small corrections of the order of magnitude given by (37) and (38). For if we neglect the terms containing the factor κ in (33), we get equation (10). From there we may pass to Hamilton's function (8) and to the corresponding Hamilton-Jacobi Equation. Noting that in the present case the latter is integrable by separation of variables, according to Levi-Civita's criterion,[36] the quantization conditions of Sommerfeld-Wilson may be applied to find the expression of the total quantized energy which lies at the base of the theory of fine structure and its applications. From a relativistic standpoint, it must be emphasized that these quantization conditions appear wholly artificial, as has been already pointed out by other authors and discussed by the present writer at another occasion.

[36] Cf. M. Sandoval Vallarta, Journal of Mathematics and Physics, Vol. 3, p. 108, 1924.

August, 1924.

(Contribution from the Research Laboratory of Physical Chemistry, Massachusetts Institute of Technology, No. 163.)

THE REPRESENTATION OF THE HABER EQUILIBRIUM DATA BY AN EQUATION[1]

By Louis J. Gillespie

1. **Introduction.** Not long ago two papers were published from the Fixed Nitrogen Research Laboratory, United States Department of Agriculture, on the Haber equilibrium, the equilibrium between nitrogen, hydrogen, and ammonia. The first paper was published by Larson and Dodge[2], and dealt with pressures from 10 to 100 atmospheres; the second, by Larson[3], dealt with pressures from 300 to 1,000 atmospheres. These data prove definitely that the equilibrium constant in terms of "partial pressures," K_p, which for ideal gases is a function only of the temperature, is in this case a (smaller) function of the pressure also.

Over a year ago the writer made an unsuccessful attempt to correlate these data thermodynamically with the aid of various hypotheses relating to mixtures of gases[4]. For this purpose the data were first smoothed for temperature. As a preliminary to another such attempt the writer has recently carried out a smoothing of the data for both temperature (T) and pressure (p) by a non-thermodynamic method, and the quite unexpected results of this purely empirical smoothing are here reported. Table I gives the values of K_p used for the present calculations.

The problem is to pass a surface near the points plotted in space using suitable functions of K_p, T, and p as coördinates. The form of the surface and the values of the coefficients must be

[1] A brief summary of the results of this paper appeared in the Proceedings of the National Academy of Sciences, **11**, 73 (1925).

[2] Jour. Am. Chem. Soc., **45**, 2918 (1923).

[3] Jour. Am. Chem. Soc., **46**, 367 (1924).

[4] Dr. Larson very kindly gave us these data in advance of publication. The present smoothing was however begun eight months after the second paper of Dr. Larson was published.

TABLE I. Observed Equilibrium Constants of Fixed Nitrogen Research Laboratory

Pressure in Atmospheres

Temperature, °C.	10	30	50	100	300	600	1,000
325	0.0401	.....	.....	.....	.....	.....	.....
350	.0266	0.0273	0.0278	.....	.....	.....	.....
375	.0181	.0184	.0186	0.0202	.....	.....	.....
400	.0129	.0129	.0130	.0137	.....	.....	.....
425	.00919	.00919	.00932	.00987	.....	.....	.....
450	.00659	.00676	.00690	.00725	0.00884	0.01294	0.02328
475	.00516	.00515	.00513	.00532	.00674	.00895	.01493
500	.00381	.00386	.00388	.00402	.00498	.00651	.....

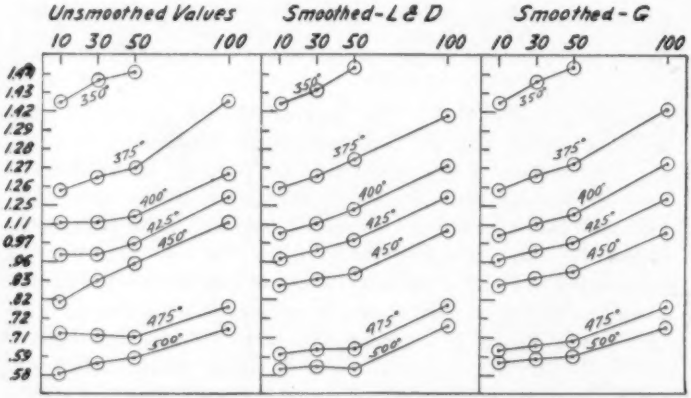
found at the same time. This requires that the element of judgment be added, and the success attained depends inversely on the arbitrariness of the judgments introduced. When there is only one independent variable, experience has shown that inflexions are presumptively to be avoided, and not to be represented unless the data strongly indicate them. In the same way we must avoid wrinkles or sudden changes of curvature when passing a surface near experimental points, unless the data strongly confirm them. Given this presumption, the problem may become practically definite in special cases in which various critics will agree that little arbitrariness of judgment is required. In such a case interpolated values given by the surface should be nearer the unknown true values than were the original unsmoothed data.

There is a natural inclination to judge the success of the smoothing by the size of the residual differences, between observed and calculated values. This inclination is stronger when we forget that precision of data is often better than the definition of the subject material or of the experimental circumstances. As a matter of fact, smoothing for several variables is a valuable method of determining in a given case whether precision exceeds definition or not.

2. **Treatment of the Low Pressure Data of Larson and Dodge.** In order to decide, with respect to which variable first to smooth, since the forms of the functions were not known, the original

values of log K_p (after addition of 3 to avoid minus signs) were plotted as a function of the pressure, from 10 to 100 atmospheres at constant temperature, with the result shown in the left-hand drawing of Fig. 1. Any decision upon a suitable smoothing function for these isotherms would seem arbitrary. When, however, the original values of log K_p were plotted as a function of $1/T$ at constant pressure, the data were found to be very consistent,

Fig. 1: Values of Log K_p +3 as Function of the Pressure (10-100 Atm.).



The sudden jumps in the values of the ordinates are due to displacing the zero of the scale for each isotherm. The isotherms are, however, plotted to the same scale. The same applies to Fig. 2.

the isobars all obviously being straight, or so nearly so that no curvature could be established. The slopes of these isobars show no detectable sudden changes from 10 to 100 atmospheres. The plot is given in Fig. 2. The differences between the observed values and those calculated from the best straight lines will be included in Table III and the case can critically be judged by inspection of these differences as to magnitude and sign.

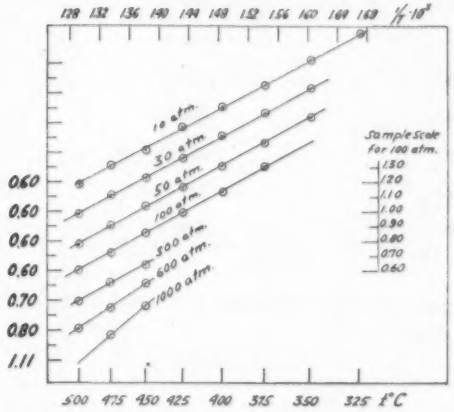
From an empirical standpoint there is a minimum of arbitrariness, therefore, in writing for the smoothing function at constant pressure:

$$\log K_p = s\ 1/T - b \qquad (1)$$

The values of s and b were determined for each pressure by the method of least squares and without weighting, to minimize personal bias, and the probable errors of s and b

$$(0.67\sqrt{(vv)/(m-2)}p_x)$$

Fig. 2: Unsmoothed Values of Log K_p +3 as Function of the Reciprocal Temperature

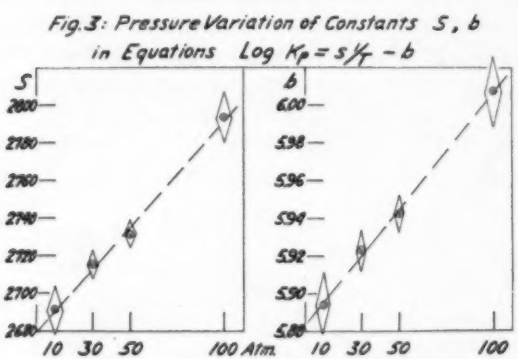


were also determined. In all the calculations $T=t+273.1$, and common logarithms are used.

The values of log K_p calculated from these values of s and b are plotted in the right-hand drawing of Fig. 1. The curves in the middle drawing give the corresponding values calculated from the smoothing equations given by Larson and Dodge. It is evident that smoothing for temperature by either their five-constant equations or the equations (1) improves the appearance of the pressure variations.

These values of s and b in equations (1) were then plotted as functions of the pressure, with the result shown in Fig. 3. The vertical diameters of the diamonds are equal to the probable errors. The broken straight lines are those calculated by the method of

least squares, weighting the points with weights equal to the reciprocals of the probable errors. The effect of the weighting is very slight. The points, of course, cannot lie exactly on the lines, as the data have not previously been smoothed for pressure.



There appears to be but little of the arbitrary[8] in writing for the parameters

$$s = Bp + A \tag{2}$$

and

$$b = Dp + C. \tag{3}$$

Combining these results we get

$$\log K_p = (A + Bp)1/T - (C + Dp) \tag{4}$$

which for this low pressure work becomes

$$\log K_p = (2679.35 + 1.1184\ p)\ 1/T - (5.8833 + 0.001232\ p) \tag{5}$$

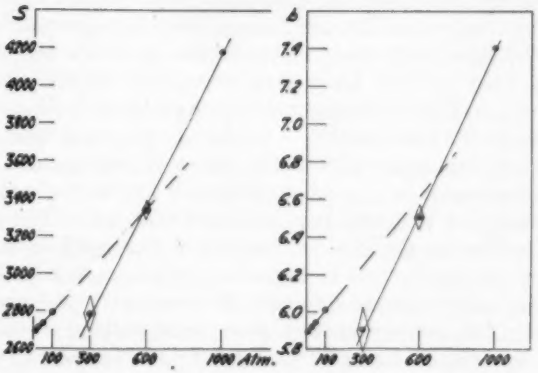
3. **Treatment of the Whole Body of Data together with Haber's Best Data.** Similar calculations were made, using the high pressure data of Larson. The case for the straightness of the isobars, though not so strong, is good. There are fewer points, in no case

[8] With regard to the discrepancy next to be pointed out the reader will observe that this discrepancy will only be increased if we pass, instead of straight lines in Fig. 3, the alternative curves with inflexion.

more than three on the isobar, and there is no consistent curvature, as may be seen later by inspection of Table III.

When the slopes and intercepts of the best isobars are plotted as functions of the pressure, together with those for the low pressure data on the same scale, the striking picture shown in Fig. 4

Fig. 4: Pressure Variation of Constants S, b in the Two-Constant Equations



results. The data are discontinuous, the point of discontinuity lying between the two sets of data. The diamonds show again the probable errors, which have naturally for the high pressures but slight value. The probable errors for the low pressure data are so small as to lie within the circles. It is not possible in either case to draw a curve without a physical discontinuity near the points within the probable errors. If the lines do not remain well within the probable errors it is found in this case that the data will be badly represented by the lines. It would be an arguable question whether this consideration ought not to be entirely neglected and (very arbitrary) lines without inflexion drawn near all the points, but for the facts to be discussed in the next three paragraphs.

In the low pressure work many more points were measured than in the high pressure work, and equilibrium was approached

from both sides, thus at least doubling the weight to be given to the low pressure points. Thus the discrepancy cannot cast doubt on the low pressure data, and it is unnecessary to cast doubt on the precision of the high pressure data. There is little or no question as to the exact composition of the original gas delivered to the catalyzer in the low pressure work. In the high pressure work, however, the exact composition is not stated. It is, however, necessary to have a value in order to calculate K_p from the equilibrium percentage of ammonia. Through the courtesy of Doctors Larson and Dodge I learn that this composition was taken in this calculation to be the same as in the low pressure work. They pointed out in their paper that "a 10% variation in the 3:1 hydrogen-nitrogen ratio produces an error of only 0.2% in the equilibrium constant." I wish to point out that such a statement can apply only to the above calculation, and not to any questions of fact, and the statement does not contradict my opinion that a 10% variation in the 3:1 ratio will in fact produce an experimental variation of much more than 0.2% in the equilibrium constant. This is supported by comparison of Haber's data for 30.09 atmospheres with the later work of Larson and Dodge. For the comparison, some extrapolation is necessary. Least squaring of Haber's best data[6] for 30.09 atmospheres gives the values for s and b, 2880.36 and 6.1264, which may be compared with the values in equation (5). A difference of 0.09 atmosphere makes only a microscopic difference in the constants s and b, and does not here affect the fourth decimal place in log K_p. The fairness of the form of equation (1) as applied to Haber's results may be judged by the deviations: 1,000 times calculated minus observed log K_p; given in Table II.

TABLE II. Representation of Haber's 30 atmosphere data by the equation: log $K_p = 2880.36\ 1/T - 6.1264$.

Temperature, °C.	561	620	631	704	710
10^3 (calc.-obs.) log K_p	−0.7	−2.0	+2.6	+4.3	−6.9
Temperature, °C.	722	812	914	952	
10^3 (calc.-obs.) log K_p	+3.0	−0.3	+1.1	−1.4	

[6] Haber. Ztschr. f. Elektrochem., **20**, 597 (1914). The H. and L. figures were not used.

The greatest deviation, 6.9, amounts to 1.6% error in K_p itself, and it is obvious that these data are of practically the same degree of consistency as those of Larson and Dodge at similar pressures (see Table III). It appears justified to extrapolate Haber's values by means of this equation to 500° C., which gives $\log K_p+3=0.5993$. Comparing this with the values 0.5866, 0.5887, and 0.5888, which are in turn (1) the L. and D. measured value, (2) my temperature-smoothed value from L. and D., and (3) the T and p smoothed value from equation (5); there is seen to be a difference in the logarithms of 0.0127 or 0.0106. The latter difference corresponds to a difference of 2.5% of K_p itself, which is outside the apparent experimental error of either set of data. To examine whether this asserted difference is dependent upon the form of function used for the temperature smoothing we may compare the percentage of ammonia calculated by Larson and Dodge for 500° and 30 atmospheres with the value calculated by Haber. These values are given by Larson and Dodge as 3.49 and 3.62% ammonia, respectively. In the calculation of these values, a Nernst heat-theorem equation was used in each case. These values lead to values of K_p which differ by 4%, Haber's value again being the greater, so that the thermodynamic smoothing does not fail to reveal the discrepancy.

It is hardly permissible to attribute this difference to experimental error, in view of the great care and labor expended by both German and American investigators.[7] It is justifiable to accept the findings as facts, and to conclude that the 2.5% variation in K_p is due to the known 6% variation in the 3:1 hydrogen-nitrogen ratio. This ratio was 3.24 in the Larson and Dodge work and 3.06 in Haber's work. The concomitant small variation in the small argon-hydrogen ratio is presumably of little importance.

In view of the very good consistency which the high pressure data of Larson exhibit when taken alone, a separate equation (6) was therefore derived for the high pressure data.

$$\log K_p=(2172.26+1.99082\ p)\ 1/T-(5.2405+0.0021550\ p) \quad (6)$$

[7] The experiments of Greenwood (Ztschr. f. Elektrochem, **21**, 241 (1915) do not have to be considered as a part of Haber's equilibrium data, as they were not planned to furnish K_p values, but to determine the practical efficiency of a catalyzer. Greenwood did not even calculate K_p values.

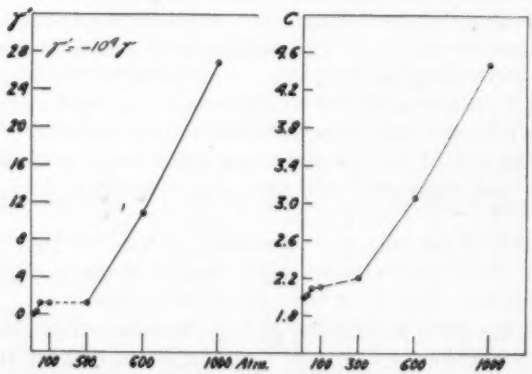
When equations (5) and (6) are both extrapolated to 200 atmospheres and the calculated K_p values are compared, the discrepancy in K_p is only 5.6% at 450° and −1.7% at 500°. Such differences might reasonably be accounted for by an assumed variation of 10% in the hydrogen-nitrogen ratio.

In a recent article[8] it has been pointed out that in a case like the present one, where K_p varies with pressure, it must also vary with the composition, and we should expect only a small effect corresponding to the small variation with the pressure. Such statements cannot be proved by thermodynamics alone, because no rigorous equations containing K_p exist, the function K_p having been introduced by means of the ideal gas laws. Such statements, once made, will, however, be regarded as obvious.

It should be emphasized that the discrepancy is a second order effect, since intentional variations of composition were not made, and it cannot be perceived until an attempt is made to smooth simultaneously for temperature and pressure.

The discrepancy is also revealed when the five-constant equation from the Nernst heat-theorem is used for the temperature smoothing, as is shown by Fig. 5.[9] This shows as functions of the

Fig. 5: Pressure Variation of Constants γ, C in the Five-Constant Equations



[8] Gillespie, Jour. Am. Chem. Soc., **47**, 305 (1925).
[9] The second abscissa should be 300 instead of 500.

pressure the two constants calculated by Larson and Dodge and by Larson. The other three constants are not varied with the pressure. In the writer's opinion, straight lines cannot be passed sufficiently near the low pressure data, disregarding the high pressure data, to represent the data within the experimental error. Therefore we are at a loss to know how to proceed, when we smooth for temperature in this way. In general, smoothing by the empirical two-constant equation appears better than by the five-constant equation, but the superiority does not appear until we begin to smooth for both variables. Haber noted a similar fact in discussing the pressure variation from 1 to 30 atmospheres[10], and noted that a smaller value of Q (our s) must be used at 1 atmosphere. He seems to have considered this latter fact as an objection, considering the two-constant equation as a mere approximation of the more theoretical five-constant equation. A theoretical advantage may, however, be illusory in a case where the ideal gas laws are now definitely known to be failing.

It is in fact noteworthy that the five-constant equation is no worse in this case, because similar equations have many times been applied to simpler, univariant, equilibria, and almost invariably the believers in the equation have been forced to conclude that the data are erroneous. This was without necessity of smoothing for a second variable. In one of these cases Keyes and Hara[11] recently proved the older data correct, and the equation wrong.

The Gibbs phase rule leads to the expectation that K_p, like any other internal variable of the system, be in this case a function of four variables; for instance, the temperature, pressure, and the ratios $H:N$ and $Ar:N$[12]. These ratios may refer, for convenience in industrial application, to the original gas mixture delivered to the catalyst, providing arbitrary changes are not subsequently made. In the present case the ratio $Ar:N$ was practically constant, but it has not been so in all the earlier experimentation.

[10] Ztschr. f. Elektrochem., **21**, 89 (1915). Especially p. 105.
[11] Jour. Am. Chem. Soc., **44**, 479 (1922).
[12] When the gases are ideal, the ratios disappear as variables, since they are incorporated into the function K_p in just the proper manner.

4. **Representation of the Data by the Equation.** It remains to compare the original unsmoothed data (*A*), with values calculated from the individual equations (1); (*B*), with values calculated from the individual equations given by Larson and Dodge and by Larson; and (*C*), with values calculated from equation (5) up to 100 atmospheres and from equation (6) from 300 to 1,000 atmospheres. In Table III are plotted, after multiplication by 1,000, the differences: calculated log K_p minus observed log K_p for all the data. The two mean differences, with, and without

TABLE III. *The differences, calculated minus observed log K_p, from A, K_f smoothing; B, L. & D. smoothing; C, K_f and p smoothing.*

	10			30			50			100			300			600			1000 Atm.		
$t°$	A	B	C	A	B	C	A	B	C	A	B	C	A	B	C	A	B	C	A	B	C
325	+25	-08	-01																		
350	+02	-05	-65	-19	-50	-26	-26	-09	+09												
375	+08	+15	-18	+13	+04	+09	+28	+53	+61	-27	-76	-51									
400	-64	-51	-89	00	00	-03	+19	+40	+50	+59	+52	+37									
425	-23	-14	-47	+28	+29	+26	+11	+19	+40	-03	00	-14									
450	+87	+88	+62	+15	+11	+15	-38	-51	-12	-49	-35	-69	+36	+53	-35	-19	-33	+185	Ø	+03	-54
475	-94	-109	-118	-58	-75	-58	-13	-53	+12	+05	+19	-14	-73	-79	-137	+40	+38	+150	Ø	+03	-48
500	+60	+26	+35	+21	-13	+22	+20	-54	+44	+15	+22	-03	+38	+23	-18	-20	-33	+78	(No data)		
Mean	0	-07	-30	0	-13	-02	0	-08	-29	0	-03	-16	0	-09	-63	0	-09	+111	Ø	+03	-51
Mean	45	90	54	22	26	20	22	40	33	26	34	31	99	95	63	26	35	111	Ø	03	51

Values are multiplied by 1000.

5. corresponds to 12% in K_p
10. " " 23 " " "
15. " " 35 " " "

respect to sign, are also given[13]. Inspection of this table shows that the two-constant equations represent the data as well as, or slightly better than, the five-constant equations. The three constants which are not functions of the pressure possibly interfere somewhat with the success, from an empirical standpoint, as the means with respect to sign show a small consistent deviation. The remarkable consistency of the low pressure data, 10–100 atmospheres, is demonstrated by the fact that the surface (5) passes practically as near the points (disregarding the sign of

[13] The first mean under 50 atmospheres, *C*, should be +2.9. Several decimal points have been lost in the cut, but they are obvious.

the deviations) as do the individual lines determined without pressure smoothing. The representation of the data from 300 to 1,000 atmospheres by the surface (6) is reasonably good, aside from the consideration that the ratio $H:N$ may have changed between points to a certain extent[14].

Summary

1. The equation:

$$\log K_p = (2679.35 + 1.1184\ p)\ 1/T - (5.8833 + 0.001232\ p)$$

represents within the experimental error the equilibrium constant values determined by the Fixed Nitrogen Research Laboratory over the temperature range about 325° to 500°C and from 100 atmospheres to the lowest pressures. The equation is serviceable for calculating "K_p" exactly for an equilibrium mixture resulting from the use of a reacting mixture made from atmospheric nitrogen and having a hydrogen-nitrogen ratio equal to 3.24.

2. The equation:

$$\log K_p = (2172.26 + 1.99082\ p)\ 1/T - (5.2405 + 0.0021550\ p)$$

represents well the high pressure data of the same laboratory, but not so many decimal places are serviceable in this case, because the hydrogen-nitrogen ratio is less definitely known.

3. An attempt to construct one equation for both high and low pressure data fails because of a discontinuity, which is too small to be shown by plotting of isobars, but is strikingly exhibited when the pressure variations of the parameters of the isobars are

[14] If values of the equilibrium percentage of ammonia are desired, the calculation can be made for the low pressure data, and must be made for the high pressure data, by means of the approximate equation: $K_p = 3.083\ a/[(0.997 - a)^2 p]$, since this was used in the calculation of the K_p values. I am indebted to Drs. Larson and Dodge for this equation. Since a certain error is involved in the use of this equation when the volume fraction of ammonia, a, is large, I have made a separate study of the high pressure data, using exactly calculated values of K_p. The same conclusions follow. The representation of the data by the surface, which now has slightly different coefficients, is not quite so good, but nearly so. The important discrepancy is in no way lessened. It was obviously unnecessary to repeat the low pressure calculations, since the error becomes much smaller.

examined. A similar discontinuity exists between the given data at 30 atmospheres and Haber's data at the same pressure. Both discontinuities are believed to be the result, not of experimental error, but of variation in the composition of the gas mixture.

4. Evidence is therefore found in the data for the Haber equilibrium that the function K_p is not strictly constant with respect to any variation whatsoever of the internal variables, the number of independent variables being given by the phase rule.

5. An important experimental task, therefore, remains; to determine, at least around the stoichiometrical composition, the variation of K_p with the composition. At high pressures the maximal yield of ammonia may occur at a hydrogen-nitrogen ratio different from three.

THE ELECTRIC CURRENTS IN A NETWORK[1]

By PHILIP FRANKLIN

If electric currents are flowing in a network of conductors, due to constant *E.M.F.*'s introduced at various places in the network, the currents in the steady state depend only on the magnitude of the *E.M.F.*'s, the magnitude of the resistances of the elements of the network, and its topology. The explicit solution for the currents in terms of these elements was given by Kirchoff[2] in a paper which was essentially the first contribution to the study of the analysis situs of the linear graph, and may be stated as follows:

In a network of wires (a wire extending between two points where more than two wires meet, two wires alone coming together at a point being regarded as a single one), whose ith wire has resistance R_i and *E.M.F.*'s F_i inserted in it.

The current in the jth wire, I_j, is a fraction whose denominator is the sum of all possible products of μ R_i's such that the corresponding wires, if removed, would leave no circuits in the network; and whose numerator is the sum of all products of $\mu-1$ R_i's such that the corresponding wires, if removed, would leave a single circuit containing the jth wire, each such product being multiplied by the E.M.F.'s (F_i's) in this circuit taken with plus or minus signs according as they are or are not in the direction of this circuit, as determined by the direction taken as standard for the jth wires.

As an example, if the *E.M.F.*'s in the wires in Fig. 1 are in the direction indicated by the arrows, the current in the wire numbered 2 in the direction of the arrow is:

$$I_2=\frac{R_1F_2-R_1F_3+R_3F_2-R_3F_1}{R_1R_2+R_1R_3+R_2R_3}$$

The object of this note is to give an alternative proof of Kirchoff's result, which is somewhat shorter than his and also brings to light the mathematical nature of the result. We shall make use of the known machinery for treating the linear graph, for example

[1] Presented to the American Mathematical Society, October 25, 1924.
[2] Annalen der Physik und Chemie. 1847. p. 498.

as given by Professor Veblen,[3] whose terminology we shall here follow. We regard our network of conductors as an oriented linear graph, whose a_0 0-cells are the points where more than two conductors meet, or where a single one ends, and whose a_1

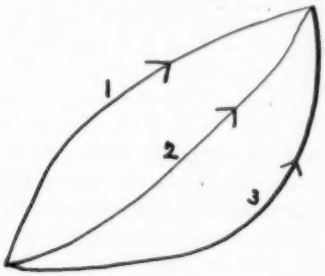


Fig. 1

1-cells are the conductors themselves (regarding as a single conductor two which come together at a point by themselves), the orientation of the cells being determined arbitrarily. We denote by I_j and F_j the current and *E.M.F.* respectively in the positive direction and by R_j the resistance in the jth 1-cell. I_j and F_j are thus signed quantities. The equations for determining the currents I_j are then:

$$(1) \qquad \sum_j \epsilon_{ij} I_j = 0, \qquad (i = 1, 2 \ldots a_0)$$

where ϵ_{ij} are the elements of the incidence matrix E_1[4], which states that the total amount of current entering the ith -cell is zero, and

$$(2) \qquad \sum_j x_j^i R_j I_j = \sum_j x_j^i F_j \qquad \text{where } \sum_j \epsilon_{ij} x_j^i = 0,$$

where the x_j^i are 0, $+1$ or -1, and the second condition insures that the set of 1-cells $(x_1^i \ldots x_j^i \ldots x_{a_1}^i)$ form an oriented circuit, and consequently (2) states that the back *E.M.F.* equals the applied *E. M. F.* in the ith circuit. Since the rank of $E_1 = \| \epsilon_{ij} \|$ is $a_0 - 1$, there are $m = a_0 - 1$ independent equations of type (1) and $\mu = a_1 - a_0 + 1$ such of type (2), corresponding to the indepen-

[3] O. Veblen, The Cambridge Colloquium Lectures on Analysis Situs, New York, 1922. pp. 23–28.

[4] lc. p. 26.

dent sets (x_j^i). Thus there are in all a_1 independent equations, which uniquely determine the a_1 I_j's.

The value of I_j determined by these equations will be D_j/D where D is the determinant of the system, with first m rows a_0-1 rows of E_1, and last μ rows the coefficients of the left members of the equations (2) used. D_j is obtained from D by replacing the elements in the jth column by zeros in the first m rows, and by the right members of the appropriate equations (2) in the last μ rows. This determinant solution is well known[8] but the explicit form of Kirchoff seems to have been overlooked.

Before proceeding to the evaluation of D_j and D, we shall prove a theorem concerning a determinant closely related to the matrix E_1, which will be useful in the evaluation. We begin with the

Lemma. In any matrix E_1 (i.e., one with two non-zero elements in each column, one $+1$ and one -1) of a_0 rows, a sub-set (not including all) of the rows can be found whose sum has zeros in $k=a_0-2$ pre-assigned places.

Let $i_1 \ldots i_k$ be the pre-assigned places. If the first row has a non-zero element in one of these places (if not, it alone is the subset required), say i_1, we add the row having the other non-zero element in the i_1th column to it, giving a zero in this place. If the sum has a non-zero element in the i_2th place, we proceed similarly with this column. This process may be continued, as the complementary element required will always be in a row not previously used, and a place once cleared will remain cleared. Furthermore the set obtained will contain at most $k+1=a_0-1$ rows and consequently will not contain all the rows. Since the sum of all the rows is zero, the complementary set may also be taken as the required subset.

We next prove the

Theorem. A determinant whose first $m=a_0-1$ rows are obtained from a matrix E_1 by deleting one row, and whose last $\mu=a_1-a_0+1$ are a set of independent solutions of the equations $\sum_j \epsilon_{ij}x_j=0$ $(E_1=||\epsilon_{ij}||)$ has all the non-zero terms in its expansion with

[8] vide, *e.g.*, Jeans, Electricity and Magnetism, pp. 311–312. Livens, Electricity, pp. 294–296.

reference to minors obtained from the first m rows equal to one another. These terms are obtained from minors which correspond to a set of independent columns of E_1.

We first note that each of the terms in the expansion formed as described will consist of a product of a determinant of the μth order from the last μ rows by the complementary one of the mth order from the first m rows. The term is determined by the columns in this last determinant, and for non-zero terms these columns must correspond to a set of independent columns of E_1. Consider first two such sets which only differ in a single column, as $i_1 \ldots i_m$ and $i_1 \ldots i_{m-1}\, i_{m+1}$. We wish to prove the corresponding terms equal. These terms will not be affected if we replace one of the first m rows of our determinant by a sum of it and some of the other rows, selected from these first m. Since the m rows contain all but one of the rows of an E_1 matrix, they must contain one of the two complementary subsets found in the lemma, and accordingly we can replace one of the first m rows by a sum with zeros in the places $i_1 \ldots i_{m-1}$. The elements p_{i_m} and $p_{i_{m+1}}$ in this row (and in columns indicated by the subscripts) cannot be zero, as the two sets of i's corresponded to independent columns of E_1, and hence are $+1$ or -1. If we denote by U the determinant formed from the unchanged $m-1$ of the first m rows and indicated columns; and by X that formed from the last μ rows and indicated columns, our two terms may be written

$$(-1)^n p_{i_m} U(i_1 \ldots i_{m-1})\, X(i_{m+1} \ldots i_{m+\mu}) \quad \text{and}$$

(3)

$$-(-1)^n p_{i_{m+1}} U(i_1 \ldots i_{m-1})\, X(i_m i_{m+2} \ldots i_{m+\mu}).$$

Since the linear combinations of the elements in the first m rows of our determinant with the elements of the last μ rows as coefficients, vanished, $(\sum_j \epsilon_{ij} x_j = 0)$, and the new row was the sum of a sub-set of the first m rows, the elements of this row (p_j) satisfy the relations

$$p_{i_m} x_{i_m} + p_{i_{m+1}} x_{i_{m+1}} + \sum_{j=m+1}^{j=m+\mu} p_{i_j} x_{i_j} = 0. \tag{4}$$

Since the x's are the elements of any one of the rows of the determi-

nants X, by substituting the values of x_{i_m} obtained from (4), and making use of an elementary theorem on determinants, we find:

$$(5) \qquad X(i_m i_{m+2} \ldots i_{m+}) = -p_{i_{m+1}}/p_{i_m} X(i_{m+1} \ldots i_{m+}).$$

Since p_{i_m} and $p_{i_{m+1}}$ are ± 1 this shows that the two terms of (3) are equal.

We complete the proof by induction, showing that the equality of two terms for sets $i_1 \ldots i_m$ and $i_1 \ldots i_{m-s} i_{m+1} \ldots i_{m+s}$ which differ in s colums follows from that for two sets differing in $s-1$ columns. As the second set of columns of E_1 is a linearly independent one, all of its members cannot be dependent on the set $i_1 \ldots i_{m-1}$. If i_{m+j} is one which is not, the set $i_1 \ldots i_{m-1} i_{m+j}$ will correspond to a non-zero term which is equal to the first of our terms by the case just proved, and to the second by the hypothesis of the induction.

We now return to the evaluation of $I_j = D_j/D$. Since D only differs from the determinant of the theorem by having an R_i multiplied in for the last μ elements of the ith column, its expansion, instead of being a sum of equal constants, will be a sum of products of μ R_i's, each product having the same coefficient. Furthermore, these products will be such that if the corresponding columns of E_1 are deleted, the remaining columns will be independent, *i.e.*, will correspond to 1-cells of the graph from which no circuit can be formed, or to a tree. This shows that

The expansion of D consists of a sum of products of μ R_i's, each product having the same coefficient, and corresponding to a set of 1-cells whose removal reduces the graph to a tree.

Since D_j is obtained from D by replacing the last μ elements in the jth column by suitable linear functions of the F_i's, the right members of the equations (2), each term will contain such a linear function of the F_i's multiplied by $(\mu-1)$ R_i's and a certain coefficient. These R_i's will be such that with R_j they correspond to a set of 1-cells whose removal reduces the graph to a tree. To determine the coefficient and function of the F_i's to be multiplied into a given product $R_{i_1} \ldots R_{i_{\mu-1}}$, we note that if the 1-cells corresponding to these subscripts are removed, there will remain

only one unbroken circuit in the graph, and that it will contain the jth 1-cell. If we replace one of the last μ rows, say the pth, of the determinants D and D_j by the linear combination of the last μ rows which corresponds to this circuit (so oriented that it is positively related to the jth 1-cell, and thus making the new $x^p{}_j = +1$), we shall not change the quotient D_j/D, nor the ratio of corresponding terms of D_j or D, since the operation at most changes these by a constant factor. But, after the change, to obtain the term of D_j containing $R_{i_1} \ldots R_{i_{\mu-1}}$ we shall have to take each of these from one of the $\mu-1$ unchanged rows, since the pth row contains none of these, and as the element from the pth row the linear function of the F_i's in the jth column. The term will consequently differ from the term $R_{i_1} \ldots R_{i_{\mu-1}} R_j$ in the expansion of D by having R_j replaced by this linear function, which from the construction of the pth row will be the sum of the F_i's in the circuit not containing the 1-cells $i_1 \ldots i_{\mu-1}$, taken with the orientation of the jth 1-cell. As this shows that all the terms in D_j will involve the constant coefficient appearing in the terms of D, this coefficient may be divided out, showing that the expression for $I_j = D_j/D$, or the current in the jth 1-cell, in the positive direction, is:

A fraction whose denominator is the sum of all possible products of μ R_i's such that the corresponding 1-cells, if removed, will reduce the graph to a tree; and whose numerator is the sum of all products of $\mu-1$ R_i's such that the corresponding 1-cells if removed will leave a single circuit containing the jth 1-cell, each such product being multiplied by the E.M.F.'s (F_i's) in this circuit with plus or minus signs according as they are or are not in the direction of the circuit determined by the jth 1-cell.

DETERMINATION OF THE STRESSES IN A BEAM BY MEANS OF THE PRINCIPLE OF LEAST WORK

By Prof. William Hovgaard

Introduction. We owe to Saint-Venant the development of the theory of torsion and bending of beams. In two classic memoirs,[1] equally remarkable for their completeness and clearness, he has given a rigorous solution of this problem, comprising in particular the determination of displacements, strains and stresses in cylindrical beams fixed at one end and loaded at the other. A description and explanation of Saint-Venant's method is found in Prof. A. E. H. Love's treatise on the "Theory of Elasticity."

When the displacements are given, it is an easy matter, by means of the general equations of elastic equilibrium and by a simple differentiation to determine the forces which produce them.

The inverse problem, to determine the displacements when the forces are given, is of far more practical importance, but had not, when Saint-Venant took up the problem, been solved in a general way, on account of the difficulty of integrating the equations in which the displacements enter and of determining the functions and arbitrary constants of integration in such a manner as to satisfy the conditions in the various cases.

In order to overcome this difficulty, Saint-Venant adopted what he called the mixed or semi-inverse method, intermeiated between the two described above. He assumed as given a part of the displacements and a part of the forces and then, by a rigorous analysis, he determined what must be the remaining displacements and the remaining forces in order to be compatible with the strain relations as well as with the conditions of equilibrium and boundary conditions. In this way he succeeded in obtaining an exact and complete solution and applied it to a great number of particular cases of practical importance.

[1] B. de Saint-Venant: "La Torsion des Prismes," L'Académie des Sciences, Vol. 14, 1856. "La Flexion des Prismes," Journal de Mathématiques, par. J. Liouville, 2 séries, Vol. 1, 1856.

Saint-Venant's method is most clearly exhibited by stating his assumptions for a specific case, such as that of a cylindrical or prismatic beam under unequal bending, being fixed at one end and loaded by a vertical force at the other. He assumed as given:

1. A part of the displacements and their relations, in that he supposed the axis to be bent in a plane curve and the elongation (or contraction) of the fibers to be proportional to their distance from the axis projected on the plane of bending.

2. A part of the forces, in that he supposed that the fibers (longitudinal elemental prisms) did not exert on one another any stress normal to their direction, although they could act on each other longitudinally. Further, that no forces acted on the bounding cylindrical surface; and that the resultant force and couple acting on the free terminal section were known, while the mode of application and distribution of these forces was unknown. Finally no body forces were acting.

3. The fixation of the beam, in that he supposed the centroid of one of the terminal sections and the central plane element of that section as well as a narrow longitudinal strip in the plane of bending of the axis to be immovable.

Saint-Venant now showed that these assumptions were compatible between themselves, and next determined the still unknown displacements and stresses as well as the distribution of the external forces which acted on the free terminal section.

Clebsch, in his work on the "Theory of Elasticity of Solid Bodies"[2] solved the problem for a beam fixed at one end and loaded in any manner at the other end including bending, torsion and tension. He made the same assumptions as Saint-Venant, stated under (2) and (3), but did not include (1), which applies only to bending in one plane. He obtained a general solution for the displacements and stresses containing an unknown function, which he decomposed into four separate functions, enabling him to deal in a simple manner with each one of the various actions: tension, bending in the two principal planes and torsion.

In the solution here offered, the inverse method, referred to above, is followed. All the external forces as well as their distri-

[2] Alfred Clebsch, "Theorie der Elasticität fester Körper," 1862; translated and supplemented with notes by Saint-Venant, "Théorie de L'Elasticité des Corps Solides," 1883.

bution over the end sections are assumed to be known, while none of the displacements are given. By means of the Principle of Least Work the stresses are determined and from these the displacements can be found. By the use of this principle and the Calculus of Variation it is possible to embody all the conditions together with the expression for the internal or elastic work of deformation in one intergral, by the variation of which the unknown stresses can be determined directly. The solution, comprising all possible conditions of loading of the beam at the free end has the same generality as that obtained by Clebsch and, being based on the fundamental principle of Least Work, gives assurance that the state of stresses to which it leads, is one to which the beam must tend, even although it may not in practice conform to it near the points of application of the external forces. In other words, it gives the ideal or natural stress distribution.

The general solution for the stresses contains unknown arbitrary functions, which are determined by integration of two partial differential equations of the second order and the solution must at the same time fulfill certain boundary conditions. In order to illustrate the method, it is applied at the end of the paper to several beams of simple section.

The present paper may be regarded as an extension of my paper in the JOURNAL OF MATHEMATICS AND PHYSICS, Vol. II, No. 4, December, 1923. In that paper a solution was given by the same method as here applied for a beam bent by a pure couple, and it was shown that the ordinary formula for bending is applicable with a slight modification to non-homogeneous materials. We shall here deal only with beams of isotropic materials.

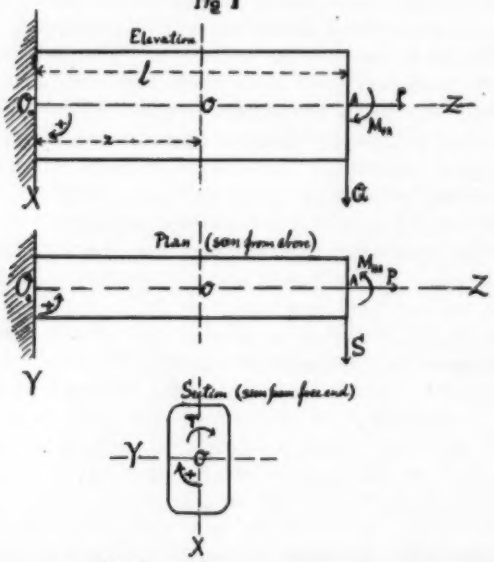
Statement of the Problem. Consider the state of stress in a cylindrical or prismatic beam of any form of section and of isotropic material, fixed at one end, while at the terminal section of the free end is applied a system of forces. No external forces act on the lateral bounding surfaces and the action of gravity is neglected.

The axis of the beam is supposed to be horizontal and a rectangular system of coördinates is placed so that the origin, O_0, is at the center of gravity of the fixed section; O_0Z coincides with the

axis, while O_0X and O_0Y coincide with the principal axes of the section. To fix ideas we take O_0X to be vertical, positive downwards and O_0Y horizontal, positive to the left when looking from the outer end towards O_0. The sign of the directions of rotation are indicated in Fig. 1. The length of the beam is $l = O_0A$.

When dealing with any section between O_0 and A we use a system of coördinates with axes in the same directions as at O_0 but with origin at O, the centroid of the section.

Fig 1



The system of forces acting on the free end represents the most general case and may be resolved into the following resultant forces and couples:

An axial force P, a vertical force Q, and a horizontal force S, all acting at the center of gravity of the terminal section at A. A torsional couple T with axis parallel to O_0A, a bending couple M_{VA} in the vertical plane and a bending couple M_{HA} in the horizontal plane. The couples produce uniform twisting or bending

moments throughout the length of the beam, but the forces Q and S produce bending moments which increase from zero at A to Ql and Sl respectively at O_0. Thus at any point distant z from O_0 we have the total bending moments:

$$\left.\begin{aligned} M_V &= M_{VA} + Q(l-z) \\ M_H &= M_{HA} - S(l-z) \end{aligned}\right\}. \tag{1}$$

It is assumed that the distribution of the external forces over the terminal section at A is such as to conform to the yet unknown stress distribution required by the Principle of Least Work. The same assumption is made for the reactions over the section at the fixed end. Thus all the forces are known or determinate and the problem is to find the stresses, which are all unknown.

It is assumed that the fixation of the beam is established in the same way as defined by Saint-Venant, making at the origin the displacements u, v, and w equal to zero, and also:

$$\left(\frac{dw}{dx}\right)_0 = 0, \qquad \left(\frac{dw}{dy}\right)_0 = 0 \qquad \text{and} \left(\frac{dv}{dx}\right)_0 = 0$$

where u, v and w are the displacements in the x, y, and z directions respectively. These conditions are necessary and sufficient to insure that the beam at the fixed end shall be prevented from bending and twisting in any direction.

As in the solutions given by Saint-Venant and Clebsch it is further assumed that there are no normal transverse stresses, that is, the longitudinal fibers, of which we may imagine the body to be made up, do not exert on each other any pressures or tensions normal to their direction.[3] Also transverse tangential tractions

[3] Saint-Venant, in his note to § 28 of Clebsch's work (referred to above), gives an elaborate discussion of this assumption and it is shown that the two normal transverse stresses can be neglected. While this is undoubtedly justified in solid beams of steel or other materials of high modulus of elasticity, it appears that in materials of low modulus such as rubber, where a rod may bend to a very sharp curve before the elastic limit is reached, normal transverse stresses of appreciable magnitude necessarily appear, directed parallel with the plane of bending towards both sides of the neutral plane and counterbalancing each other. In all thin tubular structures these forces must tend to produce collapse by transverse bending and flattening of the section, as seen when a rubber tube is bent sharply. This subject is discussed more fully in the author's work "Structural Design of Warships," 1915, pp. 61, 62.

in planes parallel with the axis are neglected. On the other hand, when the fibers move lengthwise relative to one another shear stresses are created.

There remain to be considered only three stresses: a longitudinal normal stress p; a shear stress q acting vertically in planes normal to OZ together with a corresponding stress q acting longitudinally in horizontal planes, and a shear stress s acting horizontally in planes normal to OZ with a corresponding stress s acting longitudinally in vertical planes.

It follows that the strain in direction of OZ is:

$$e_{zz} = \frac{p}{E}$$

where E is Young's modulus of elasticity, and the strain components due to shearing are:

$$\left.\begin{aligned} & e_{zx} = \frac{q}{G} \text{ and } e_{zy} = \frac{s}{G} \\ \text{Also} \quad & e_{xx} = e_{yy} = -\sigma e_{zz} = -\sigma \frac{p}{E} \end{aligned}\right\} \tag{2}$$

where σ is Poisson's ratio and G is the coefficient of rigidity. It is assumed that E and G are related by the equation.

$$G = \frac{E}{2(1+\sigma)}.$$

The Elastic Work of Deformation and the Equations of Condition.

The elastic work of deformation, expressed in terms of the stresses, takes the simple form:

$$W = \int_{x_0}^{x_1}\int_{y_0}^{y_1}\int_{0}^{l} \left\{ \frac{p^2}{2E} + \frac{q^2}{2G} + \frac{s^2}{2G} \right\} dxdydz. \tag{3}$$

The stresses p, q, and s are unknown functions of x, y, and z, but must be such as to make the elastic work a minimum while at the same time satisfying the conditions of equilibrium, the boundary conditions, and certain equations of compatability of the strains.

The equations of condition are the following:

1. For equilibrium of an elemental prism we have, since no normal forces act on the lateral faces:

$$\frac{\partial q}{\partial z}=0 \quad \text{and} \quad \frac{\partial s}{\partial z}=0 \tag{4}$$

$$\frac{\partial p}{\partial x}+\frac{\partial s}{\partial y}+\frac{\partial p}{\partial z}=0. \tag{5}$$

It follows from (4) that q and s are independent of z, whence the shear stresses are the same at any set of values (x, y) throughout the beam, *i.e.*, the tangential tractions are identically alike on every transverse section.

2. The strain components must comply with certain equations of compatibility, which, expressed in terms of the stresses according to (2), furnish the relations:[4]

$$\frac{1}{G}\cdot\frac{\partial}{\partial y}\left(\frac{\partial s}{\partial x}-\frac{\partial q}{\partial y}\right)+2\frac{\sigma}{E}\frac{\partial^2 p}{\partial z\partial x}=0 \tag{6}$$

$$\frac{1}{G}\frac{\partial}{\partial x}\left(\frac{\partial s}{\partial x}-\frac{\partial q}{\partial y}\right)-2\frac{\sigma}{E}\frac{\partial^2 p}{\partial z\partial y}=0. \tag{7}$$

3. In any section the stresses must balance the external forces P, Q, S and the moments T, M_V, M_H.

Hence at any transverse section we must have:

$$P=\int\int p\,dx\,dy \tag{8}$$

$$Q=\int\int q\,dx\,dy \tag{9}$$

$$S=\int\int s\,dx\,dy \tag{10}$$

$$T=\int\int (sx-qy)\,dx\,dy \tag{11}$$

$$M_V=M_{VA}+Q(l-z)=-\int\int px\,dx\,dy \tag{12}$$

$$M_H=M_{HA}-S(l-z)=\int\int py\,dx\,dy. \tag{13}$$

[4] A. E. H. Love, Theory of Elasticity, 1920, p. 50.

The integration in all these integrals to extend over the entire surface of the cross section.

4. At the lateral bounding surface the sum of the stress components normal to the surface must be zero since no external forces act on that surface. Let the equation to the bounding curve of a transverse section be:

$$y=\phi(x) \quad \therefore \quad \frac{dy}{dx}=\frac{d\phi}{dx}. \tag{14a}$$

Then, since p acts parallel to the surface, we need only consider q and s, and we find:

$$\frac{s}{q}=\frac{dy}{dx}=\frac{d\phi}{dx} \tag{14b}$$

at all points of the boundary.

The Variation. In forming the variational integral U we follow the method of Lagrange. The expressions on the left side of the differential equations (5), (6), and (7) are multiplied by coefficients which are functions of the independent variables and are entered under the integral sign of (3), but since there are only three indeterminate quantities we can have only two independent conditions and, hence, shall need only two arbitrary coefficients. We therefore introduce the functional coefficients λ_1, for equation (5) and λ_2 for the sum of (6) and (7). It is shown below that the latter equations are not independent of each other, but are partial differentials of one equation; moreover, the same result would have been obtained if we had introduced three λ coefficients.

Equations (8) to (13) do not contain the indeterminate functions p, q, and s on the left side, and on the right side we find double integrals. In order to incorporate them under the triple integral sign of (3) we imagine them all to be integrated with respect to z, so that we get for instance:

$$\iiint p\,dx\,dy\,dz - Pl = 0 \tag{8'}$$

$$\iiint px\,dx\,dy\,dz + M_{VA}\,l + \tfrac{1}{2}Q\beta = 0. \tag{12'}$$

Each of the expressions (8′) to (13′) is now multiplied by a constant, a, b, c, e, f, g respectively and added under the integral sign of (3), but since the terms outside the integrals, such as Pl in (8′) etc., do not contain the indeterminate functions they will disappear in the process of variation and are for the sake of briefness omitted from the expression for U given below. We have thus:

$$U=\int_{x_2}^{x_1}\int_{y_2}^{y_1}\int_{0}^{l}\left\{\frac{p^2}{2E}+\frac{q^2}{2G}+\frac{s^2}{2G}+ap+bq+cs+e(sx-qy)+fpx+gpy\right.$$
$$+\lambda_1\left(\frac{\partial q}{\partial x}+\frac{\partial s}{\partial y}+\frac{\partial p}{\partial z}\right)+\lambda_2\left(\frac{1}{G}\frac{\partial^2 s}{\partial y\partial x}-\frac{1}{G}\frac{\partial^2 q}{\partial y^2}+2\frac{\sigma}{E}\frac{\partial^2 p}{\partial z\partial x}+\frac{1}{G}\frac{\partial^2 s}{\partial x^2}\right.$$
$$\left.\left.-\frac{1}{G}\frac{\partial^2 q}{\partial x\partial y}-2\frac{\sigma}{E}\frac{\partial^2 p}{\partial z\partial y}\right)\right\}dxdydz \qquad (15)$$

In order to make U a minimum it is necessary that the variation of U shall be zero when p, q and s are made to vary independently and arbitrarily by an infinitesimal amount. The variation consists of two parts, first, the terms at the limits, three groups of double integrals depending on the variations of p, q and s and of their derivatives for such values of the independent variables as satisfy the boundary equations; second, of one triple integral depending on entirely arbitrary variations of p, q and s. We shall deal first with the triple integral and show afterwards that the terms at the limits all vanish.[8] We have

$$\delta U=\int_{x_2}^{x_1}\int_{y_2}^{y_1}\int_{0}^{l}\left\{\left[\frac{p}{E}+a+fx+gy-\frac{\partial\lambda_1}{\partial z}+\frac{2\sigma}{E}\frac{\partial^2\lambda_2}{\partial z\partial x}-\frac{2\sigma}{E}\frac{\partial^2\lambda_2}{\partial z\partial y}\right]\delta p\right.$$
$$+\left[\frac{q}{G}+b-ey-\frac{\partial\lambda_1}{\partial x}-\frac{1}{G}\frac{\partial^2\lambda_2}{\partial y^2}-\frac{1}{G}\frac{\partial^2\lambda_2}{\partial x\partial y}\right]\delta q$$
$$\left.+\left[\frac{s}{G}+c+ex-\frac{\partial\lambda_1}{\partial y}+\frac{1}{G}\frac{\partial^2\lambda_2}{\partial y\partial x}+\frac{1}{G}\frac{\partial^2\lambda_2}{\partial x^2}\right]\delta s\right\}dxdydz\ \ldots \qquad (16)$$

[8] The variation of triple integrals is fully explained in Jellett's "Calculus of Variation," 1850.

Solution for the Longitudinal Stress. Since δp, δq and δs are quite arbitrary and independent of each other, we give λ_1, and λ_2 such values as shall make the factors of δq and δs vanish and it follows that also the factor of δp must then be made equal to zero in order to make the triple integral vanish.

Equating the factors of δq and δs to zero, we obtain expressions for q and s in terms of x, y, λ_1 and λ_2, but since q and s are independent of z according to (4), the coefficients λ_1, and λ_2 can be functions of x, y only. Hence the three differentials with respect to z of these coefficients, which occur in the factor of δp must vanish and by equating the remaining terms to zero we obtain the equation:

$$p = -E(fx+gy+a) \tag{17}$$

Substituting in (8), we get:

$$P = -E\int\int(fx+gy+a)dxdy = -E(fm_y+gm_x+aA)$$

where m_y and m_x are the area moments of the section about the X-axis and Y-axis respectively through the centroid and A is the area of the section. Since OX and OY are by supposition the principal axes, m_y and m_x vanish and we get:

$$a = -\frac{P}{AE} = -\frac{p_0}{E} \tag{18}$$

where p_0 is the mean longitudinal stress in the beam due to the longitudinal force P.

Substituting the value of p from (17) in (12) and (13), we get:

$$M_{VA}+Q(l-z) = E\int\int(fx^2+gxy+ax)dxdy = EfI_y$$

$$M_{HA}-S(l-z) = -E\int\int(fxy+gy^2+ay)dxdy = -EgI_x$$

where I_y and I_x are the moments of inertia of the section about the axes OX and OY respectively, while all integrals containing odd powers of x and y vanish. Hence

$$f = \frac{M_{VA}+Q(l-z)}{EI_y} \quad \text{and} \quad g = -\frac{M_{HA}-S(l-z)}{EI_x} \tag{19}$$

Substituting these values of the constants in (17), we find:

$$p=-\frac{M_{VA}+Q(l-z)}{I_y}x+\frac{M_{HA}-S(l-z)}{I_x}y+p_0\ldots \quad (20)$$

The first term in this expression taken by itself gives the usual formula for bending in the vertical plane. It consists of one part corresponding to a uniform bending moment due to the couple M_{VA} and another part due to the vertical end load Q. The second term gives the corresponding formula for bending moments in the horizontal plane, both being of the general form:

$$p=\frac{M}{I}x. \quad (21)$$

This is the ordinary formula for bending stresses, which is thus seen to be in accordance with the Principle of Least Work not only in case of uniform bending in a circle but also for unequal bending.

The third term in (20) represents a longitudinal stress due to the axial force P. Since neither x nor y enter into this term, we may infer that for Least Work this stress must be uniformly distributed.

It is now possible to simplify the factor of λ_3 in (15):

From (20) we find:

$$\left.\begin{aligned}\frac{\partial^2 p}{\partial z\partial x}&=\frac{Q}{I_y} \quad \text{and} \quad \frac{\partial^2 p}{\partial z\partial y}=\frac{S}{I_x}\end{aligned}\right\} \quad (22)$$

whence from (6) and (7):

$$\frac{1}{G}\frac{\partial}{\partial y}\left(\frac{\partial s}{\partial x}-\frac{\partial q}{\partial y}\right)=-\frac{2\sigma}{E}\frac{Q}{I_y} \quad (6')$$

$$\frac{1}{G}\frac{\partial}{\partial x}\left(\frac{\partial s}{\partial x}-\frac{\partial q}{\partial y}\right)=\frac{2\sigma}{E}\frac{S}{I_x}. \quad (7')$$

Integrating we obtain

$$\frac{\partial s}{\partial x}-\frac{\partial q}{\partial y}=2\tau G+\frac{2\sigma G}{E}\left(\frac{Sx}{I_x}-\frac{Qy}{I_y}\right) \quad (23)$$

where τ is a constant of integration. The expression on the left side of (23) takes the place of (6) and (7) in the factor of λ_2 in equation (15), and thus the triple integral in δU becomes:

$$\int_{x_2}^{x_1}\int_{y_2}^{y_1}\int_{0}^{l}\left\{\left[\frac{p}{E}+a+fx+gy\right]\delta p+\left[\frac{q}{G}+b-ey-\frac{\partial\lambda_1}{\partial x}+\frac{\partial\lambda_2}{\partial y}\right]\delta q\right.$$
$$\left.+\left[\frac{s}{G}+c+ex-\frac{\partial\lambda_1}{\partial y}-\frac{\partial\lambda_2}{\partial x}\right]\delta s\right\}dxdydz=0. \qquad (16')$$

We have already disposed of the factor of δp and equating now the factors of δq and δs to zero we find:

$$q=G\left[\frac{\partial\lambda_1}{\partial x}-\frac{\partial\lambda_2}{\partial y}+ey-b\right], \qquad s=G\left[\frac{\partial\lambda_1}{\partial y}+\frac{\partial\lambda_2}{\partial x}-ex-c\right] \qquad (25)$$

from which:

$$\left.\begin{aligned}\frac{\partial q}{\partial x}&=G\left[\frac{\partial^2\lambda_1}{\partial x^2}-\frac{\partial^2\lambda_2}{\partial x\partial y}\right], \quad \frac{\partial s}{\partial x}=G\left[\frac{\partial^2\lambda_1}{\partial x\partial y}+\frac{\partial\lambda_2^2}{\partial x^2}-e\right]\\ \frac{\partial q}{\partial y}&=G\left[\frac{\partial^2\lambda_1}{\partial y\partial x}-\frac{\partial^2\lambda_2}{\partial y^2}+e\right], \quad \frac{\partial s}{\partial y}=G\left[\frac{\partial^2\lambda_1}{\partial y^2}+\frac{\partial^2\lambda_2}{\partial y\partial x}\right]\end{aligned}\right\} \qquad (26)$$

Equation (5) becomes:

$$\frac{\partial q}{\partial x}+\frac{\partial s}{\partial y}+\frac{Qx}{I_y}+\frac{Sy}{I_x}=0 \qquad (5')$$

from which by substitution from (26)

$$\frac{\partial^2\lambda_1}{\partial x^2}+\frac{\partial^2\lambda_1}{\partial y^2}=-\frac{1}{G}\left(\frac{Qx}{I_y}+\frac{Sy}{I_x}\right) \qquad (27)$$

From (23) and (26):

$$\frac{\partial^2\lambda_2}{\partial x^2}+\frac{\partial^2\lambda_2}{\partial y^2}=2(\tau+e)+\frac{2\sigma}{E}\left(\frac{Sx}{I_x}-\frac{Qy}{I_y}\right) \qquad (28)$$

We shall now consider the terms at the limits. The integration is extended to the space inside of a cylinder:

$$\phi(x,y)=0$$

bounded by the end sections lying in the transverse planes:

$$z=0 \quad \text{and} \quad z=l$$

The terms at the limits must vanish for such values of δp, δq, and δs as satisfy these equations, and if the integration is first effected with respect to z between the limits $z=0$ and $z=l$, we obtain three double integrals with limits defined by $\phi(x,y)=0$. The first of these integrals is:

$$\int_{x_2}^{x_1}\int_{y_2}^{y_1}\left\{\left[\lambda_1-\frac{2\sigma}{E}\frac{\partial\lambda_2}{\partial x}+\frac{2\sigma}{E}\frac{\partial\lambda_2}{\partial y}\right]\delta p\right\}_{z=0}^{z=l}dxdy$$

which cannot be made to vanish for all values of the independent variables that satisfy $\phi(x,y)=0$ except by making $\delta p=0$. We therefore make the assumption that δp is zero everywhere in the terminal sections, and thus the first of the terms at the limits will vanish.

The two other terms containing δq and δs vanish because all the terms inside the brackets are multiplied by derivatives of z with respect to x and y, all of which vanish at the cylindrical surface.

Since, therefore, all the terms at the limits disappear, the triple integral in (16′) gives the whole variation.

We have thus arrived at the general solution of the problem, comprising two partial linear differential equations of the second order (27) and (28), but the integration of these in any given case depends on the possibility of finding such expressions for λ_1, and λ_2 as will satisfy the boundary conditions. This part of the problem depends upon the form of the contour of the transverse sections and can be completely and accurately solved for a number of sections of simple form. Although the equations (27) and (28) are very different in form from those to which Saint-Venant and Clebsch arrived, they lead to the same results.

We shall first give a solution comprising only rational integral algebraical functions of x and y, which satisfy the boundary conditions for various simple sections with one or two axes of symmetry. As shown below it will be necessary in certain cases to introduce transcendental series.

We find:

$$\left.\begin{aligned}\lambda_1 &= B(y^3-3yx^2)+C(x^3-3xy^2)-\left[\frac{Sx}{I_x}+\frac{Qy}{I_y}\right]\frac{xy}{2G}\\ \lambda_2 &= (\tau+e)x^2+\frac{\sigma}{3E}\left[\frac{Sx^3}{I_x}-\frac{Qy^3}{I_y}\right].\end{aligned}\right\} \quad (29)$$

Terms of the first power and constant terms might be added, but they would not alter the form of q and s to which we arrive. Terms of a power higher than the third might also be added, but it is found that the expressions (29) are sufficient in the simple cases which are to be here considered.

Differentiate λ_1 and λ_2 and substitute in (25):

$$q=G\left[3C(x^2-y^2)-\left(6B+\frac{S}{GI_x}\right)xy-\frac{Qy^2}{EI_y}+ey-b\right] \quad (30)$$

$$s=-G\left[3B(x^2-y^2)+\left(6C+\frac{Q}{GI_y}\right)xy+\frac{Sx^2}{EI_x}-(2\tau+e)x+c\right] \quad (31)$$

These expressions are similar in form as regards x and y, as they should be in order to cover the case of symmetry about both of the principal axes. The three constants b, c, and e can be determined from (9), (10) and (11) respectively and the three constants of integration τ, B and C from the boundary conditions.

In case of simple bending in a plane of symmetry without any other action, the stresses must be the same at corresponding points with reference to the axis of symmetry. For instance, in case of a vertical end load acting alone q must be the same for $+y$ and $-y$ for a given value of x. Hence we cannot have y to the first power in the expression for q, and it follows that $B=0$ and $e=0$. Moreover $S=0$. For a horizontal end load acting alone in a plane of symmetry we must have $Q=0$, $C=0$, $\tau=0$, $e=0$.

For a beam with two axes of symmetry under a pure torsional couple and in the absence of bending there can be no terms of even power in x or y in the expressions for the stresses, whence B and C must vanish. In that case also terms containing Q and S vanish and $b=c=0$. Substituting the resulting values of q and s in (11) we have the torsional couple.

$$T=G\int\int\left[(2\tau+e)x^2-ey^2\right]dxdy=2\tau G+e(I_y-I_x) \qquad (32)$$

where the integration is to extend over the entire area of the section. If we had entered the full expressions (30) and (31) for q and s, we should still get the same result since all the integrals containing odd powers of x or y would vanish. Whenever $T=0$ we have $e=0$ and it follows that also $\tau=0$ in the cases to which (30) and (31) apply.

We have assumed that the forces at the free end and the reactions at the fixed end are distributed over the terminal sections in accordance with the stress distribution subsequently determined by the Principle of Least Work. In practice these conditions are never fulfilled, but as there is an inherent tendency of the stresses to conform to that principle, we are justified in asserting that at a certain distance from the terminals the stresses will adjust themselves in accordance therewith. Saint-Venant has shown that this adjustment actually takes place very close to the terminals (Saint-Venant's Principle).

It has been shown here that by means of the Principle of Least Work it is possible to determine the stresses when the external forces acting on the beam are known without making any *a priori* assumptions regarding the displacements as by Saint-Venant's method. Once the stresses are determined, the displacements can be found as explained in Love's "Theory of Elasticity" (Third Ed., page 338). In this process the conditions of fixation at the origin serve to determine the constants of intergration.

It is a not uncommon fallacy to believe that the ordinary formula for bending depends in all cases for its correctness on the assumption that plane cross sections of a beam remain plane and normal to the longitudinal filaments after bending. This condition is fulfilled only in pure bending where the beam takes a circular form, but as shown by Saint-Venant[8] and as brought out in the present paper, the ordinary bending formula is rigorously true also in unequal bending, provided the shear force is constant, as in a cantilever beam loaded at the end by a transverse force. In that case the cross sections do not remain plane,

[8]L'Académie des Sciences, Vol. 14, 1856, page 187.

but curve over and incline equally towards the filaments, so that the proportionality of the longitudinal stresses expressed by the ordinary formula is strictly maintained.[6]

Application of the Formulas to Cylindrical Beams of Various Sections

I. *Elliptic Cylinder under Vertical End Load*

Consider the simple case where a cantilever beam of elliptical section is loaded only by a system of vertical forces acting over the terminal section and having their resultant at the centroid of that section.

Let the equation to the transverse section be:

$$\frac{x^2}{m^2}+\frac{y^2}{n^2}=1 \tag{33}$$

In the equations (30) and (31) for the shear stresses q and s the terms containing S vanish and $B=0$, $e=0$, $c=0$. Also, since the torsional moment $T=0$, we know from (32) that $\tau=0$. Hence we find:

$$q=G\left[3C(x^2-y^2)-\frac{Qy^2}{EI_y}-b\right] \tag{30'}$$

$$s=-G\left[6C+\frac{Q}{GI_y}\right]xy \tag{31'}$$

From (9):

$$Q=\int\int q\,dx\,dy=G\left[3C(I_y-I_x)-\frac{QI_x}{EI_y}-bA\right]$$

where A is the sectional area. Let:

$$\frac{Q}{A}=q_0=\text{mean stress.}$$

$$\therefore\ b=-\frac{q_0}{G}+3C\frac{I_y-I_x}{A}-\frac{q_0I_x}{EI_y} \tag{34}$$

[6] Saint-Venant applying his semi-inverse method, says: (l.c. page 95): "... les sections du prisme *se courbent et s'inclinent toutes également* sur les fibres, car les portions de fibres qu'elles interceptent éprouvent les mêmes dilatations que celles qui seraient comprises entre des plans restés normaux."

It remains to find C and b from the boundary condition

$$\frac{s}{q}=\frac{dy}{dx}=-\frac{xn^2}{ym^2} \quad \therefore\ qxn^2+sym^2=0 \tag{35}$$

Substituting from (30′) and (31′):

$$3Cn^2x^3-3Cn^2xy^2-\frac{Qn^2}{EI_y}xy^2-bn^2x-6Cm^2xy^2-\frac{Qm^2}{GI_y}xy^2=0.$$

Since $x^2=m^2\left(1-\frac{y^2}{n^2}\right)$ at the boundary, we find after substitution and division by x:

$$(3Cm^2n^2-bn^2)=\left(9Cm^2+3Cn^2+\frac{Qn^2}{EI_y}+\frac{Qm^2}{GI_y}\right)y^2. \tag{36}$$

This equation must hold for all values of y consistent with the boundary equation, whence it is necessary that the constant terms and the factor of y^2 must vanish separately. It follows from (36) and (34) that:

$$b=3Cm^2=-\frac{q_0}{G}+3C\frac{I_y-I_x}{A}-\frac{q_0 I_x}{E I_y}.$$

Since $I_y=\frac{\pi m^3 n}{4}$ and $I_x=\frac{\pi m n^3}{4}$ we find:

$$C=-\frac{4q_0}{3m^2EG}\frac{(m^2E+n^2G)}{(3m^2+n^2)} \tag{37}$$

If we equate the factor of y^2 in (36) to zero we obtain directly the same value for C.

It follows that:

$$b=-\frac{4q_0}{EG}\frac{(m^2E+n^2G)}{(3m^2+n^2)}$$

and

$$q=\frac{4q_0}{m^2(3m^2+n^2)E}\left[(m^2E+n^2G)(m^2-x^2)-m^2(3G-E)y^2\right] \tag{38}$$

$$s=\frac{4q_0}{m^2}\left[\frac{2(m^2E+n^2G)}{(3m^2+n^2)E}-1\right]xy \tag{39}$$

In case of horizontal end load we have $C=0$ while B is retained in (30) and (31). The solution is quite analogous to that for vertical end load, only s_0 takes the place of q_0 and m and n are interchanged.

For a *circular cylinder* these equations become:

$$q=\frac{q_0}{a^2E}\left[(E+G)(a^2-x^2)-(3G-E)y^2\right] \tag{38'}$$

$$s=-\frac{2q_0(E-G)}{a^2E}xy \tag{39'}$$

At the neutral axis:

$$q=\frac{q_0}{E}\left[(E+G)-(3G-E)\frac{y^2}{a^2}\right] \quad \text{and} \quad s=0. \tag{38''}$$

According to the commonly used formula for the shearing stress at the neutral axis:

$$q=\frac{4}{3}\,q_0$$

which in this particular case differs very little from (38'').

II. *Elliptic Cylinder under Torsion*

We assume here that a pure torsional couple T acts on the terminal section. In that case $Q=S=b=c=0$ and since there can be no even powers of x and y in the expressions for q and since B and C vanish, we have:

$$q=Gey \tag{30''}$$

$$s=G(2\tau+e)x \tag{31''}$$

$$\therefore\ T=\int\int(sx-qy)dxdy=G[(2\tau+e)I_y-eI_x]$$

$$e=\frac{4T-2AGm^2\tau}{AG(m^2-n^2)}. \tag{40}$$

The boundary equation (35) gives:

$$en^2yx+(2\tau+e)m^2xy=0$$

$$\therefore\ e=-\frac{2\tau m^2}{m^2+n^2}. \tag{41}$$

From (40) and (41):

$$\tau = \frac{(m^2+n^2)}{m^2n^2AG} \tag{42}$$

and

$$e = -\frac{2T}{n^2AG}. \tag{43}$$

The constant τ is the angle of torsion per unit length.

From (30′) and (31′):

$$q = -\frac{2T}{n^2A}y = -2\tau G\frac{m^2}{m^2+n^2}y \quad \text{and} \quad s = \frac{2T}{m^2A}x = 2\tau G\frac{n^2}{m^2+n^2}x. \tag{44}$$

The resultant stress reaches its maximum as a tangential stress at the end of the short axis of the elliptical contour and is: $\frac{2T}{An}$.

For an elliptic tube the walls of which are bounded by two similar ellipses of semi axes m, n and km, kn, where $k<1$, we find by proceeding in the same way as above:

$$\tau = \frac{(m^2+n^2)T}{m^2n^2(1-k^4)AG} \tag{42′}$$

$$q = -\frac{2T}{n^2(1-k^4)A}y \qquad s = \frac{2T}{m^2(1-k^4)A}x \tag{44′}$$

where A is the area of the full section.

For a beam of full circular section we have

$$\tau = -e = \frac{T}{I_0G}, \qquad q = -\frac{T}{I_0}y, \qquad s = \frac{T}{I_0}x \tag{45}$$

where I_0 is the polar moment of inertia of the transverse section.

The tangential stress at any radius r is found by combining q and s to be:

$$\frac{T}{I_0}r.$$

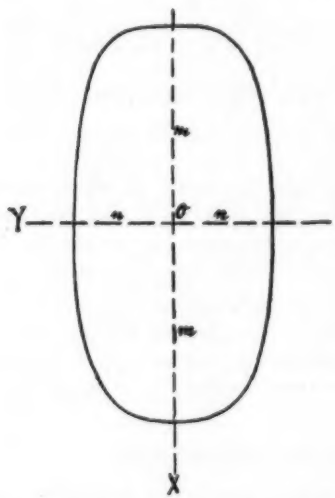
III. *Cylindrical Beam of Quasi-Elliptical Section under Vertical End Load.*

Let the equation to the contour of the section be:

$$\frac{x^2}{m^2}+\frac{y^4}{n^4}=1. \tag{46}$$

which is more nearly rectangular than an ordinary ellipse. Fig. 2 represents the case where $m=2n$. The symbol I is used instead of I_y, as I_x does not occur.

Fig. 2.



The beam is loaded only by a vertical end load Q.
Since:

$$\frac{dy}{dx}=-\frac{n^4x}{2m^2y^3}$$

the boundary condition becomes:

$$2sm^2y^3+qn^4x=0 \tag{47}$$

Since the section is symmetrical, we have as in Case I: $B=0$,

and we may use the values of q and s given in (30′) and (31′) Substituting in (47):

$$-12Cm^2y^4x-2m^2\frac{Q}{GI}y^4x+3Cn^4x^3-\left(3C+\frac{Q}{EI}\right)n^4y^2x-bn^4x=0.$$

Divide by x, substitute $x^2=m^2\left(1-\frac{y^4}{n^4}\right)$ and equate sum of constant terms to zero:

$$\therefore\ b=3Cm^2.$$

From the remaining terms we obtain:

$$C=-\frac{Q}{3GEI}\,\frac{2m^2Ey^2+n^4G}{5m^2y^2+n^4}.$$

Now let $\sigma=\frac{1}{4}$, whence $G=\frac{2}{5}E$ and we find:

$$C=-\frac{Q}{3EI}\qquad\therefore\ b=-\frac{Qm^2}{EI}.$$

Substituting these values in (30′) and (31′) we find:

$$q=\frac{2Q}{5I}(m^2-x^2)\quad\text{and}\quad s=-\frac{Q}{5I}xy. \tag{48}$$

This solution is simpler than that for a circular cylinder due to the remarkable fact that the vertical shear stress is independent of y as in the commonly used formula for bending. Also the moment of inertia I_x is absent. The maximum vertical shearing stress is found in the neutral axis and is

$$q_{\max.}=\frac{2Q}{5I}m^2.$$

The horizontal stress s for any given value of x has its greatest values at the contour, and reaches its maximum when $y=\frac{n}{\sqrt[4]{3}}$ or at about $y=\frac{3}{4}n$. Here, very approximately:

$s_{\max.}=-\frac{Q}{8I}mn$. If $m=2n$ we find the numerical ratio between the maximum stresses.

$$\frac{s_{\max.}}{q_{\max.}}=\frac{5}{32}$$

showing that the transverse stress reaches a magnitude of some-

what less than one-sixth of the vertical stress. If on the other hand $n=2m$, that is, the case of a flat shallow beam, we find $s_{max.}=\frac{8}{9}\,q_{max.}$, showing that in such a case and in beams still more flat, it is not safe to disregard the transverse shearing stress. This, of course, is well known, but is perhaps sometimes overlooked in practice.

IV. *Beam of Equilateral Triangular Section. Pure Torsion.*

Let the length of the sides of the triangular section be a and the height h. We have then

$$\frac{a^2}{h^2}=\frac{4}{3}, \qquad \frac{a}{h}=\frac{2}{\sqrt{3}}.$$

The equations to the sides of the triangle are:

$$\left.\begin{aligned} y&=\frac{a}{3}+\frac{ax}{2h} & \therefore\ \frac{dy}{dx}&=+\frac{a}{2h}\\ -y&=\frac{a}{3}+\frac{ax}{2h} & \therefore\ \frac{dy}{dx}&=-\frac{a}{2h}\\ x&=\frac{h}{3} & \frac{dy}{dx}&=\infty \end{aligned}\right\}. \tag{49}$$

In (30) and (31) we have: $b=c=Q=S=0$ and we also make $C=0$, retaining the arbitrary constant B.

$$\therefore\ q=-G(6Bxy-ey) \tag{30'}$$

$$s=-G\left[3B(x^2-y^2)-(2\tau+e)x\right]. \tag{31'}$$

For the sloping sides the boundary condition gives:

$$\pm\frac{a}{2h}=\frac{s}{q}$$

$$\therefore\ 3B(x^2-y^2)-(2\tau+e)x=\pm\frac{3a}{h}Bxy\mp\frac{ae}{2h}y.$$

Substitute for y from (49): $y=\pm\left[\frac{a}{3}+\frac{ax}{2h}\right]$

$$3Bx^2-3B\left[\frac{a^2}{9}+\frac{a^2x^2}{4h^2}+\frac{a^2x}{3h}\right]-(2\tau+e)x=\frac{3aB}{h}\left[\frac{ax}{3}+\frac{ax^2}{2h}\right]-\frac{ae}{2h}\left[\frac{a}{3}+\frac{ax}{2h}\right].$$

Equating constant terms and terms in x separately to zero, we find:

$$B=\frac{e}{2h}, \qquad e=-\tau, \qquad \therefore\ B=-\frac{\tau}{2h}$$

$$q=\frac{G\tau}{h}(3xy-hy), \qquad s=\frac{G\tau}{2h}\left[3(x^2-y^2)+2hx\right] \tag{50}$$

which satisfy the boundary conditions on the sloping sides. For the lower side of the triangle we have $x=\frac{h}{3}$, which gives $q=0$, as it should do. The terms in x^2 give an identity.

The torsional angle τ is found from (11), which gives:

$$T=\frac{G\tau ah^3}{30} \qquad \text{or } \tau=\frac{80T}{a^4G\sqrt{3}}$$

$$\left.\begin{aligned} \therefore\ q&=\frac{160T}{3a^5}(3xy-hy) \\ s&=\frac{80T}{3a^5}[3(x^2-y^2)+2hx] \end{aligned}\right\}. \tag{51}$$

The maximum stress exists at the middle of the sides of the triangle and is readily determined for the lower side:

$$s_{\max.}=\frac{20T}{a^3}. \tag{52}$$

V. *Beam of Rectangular Section. Vertical End Load.*

This problem cannot be solved by means of simple integral algebraic functions of x and y. It is necessary in the integration of the differential equations (27) and (28) to resort to transcendental series. The problem was solved by Saint-Venant[7] and adopting his method in determining the unknown functions we obtain the following expressions for λ_1, and λ_2, where a and b are the half sides of the rectangle forming the contour of the section, n is an integral number. The symbol I is used for I_y.

[7] Liouville's Journ. de Math. 1856, pp. 156 *et seq.*

$$\left.\begin{aligned}\lambda_1 &= -\frac{Qxy^2}{2IG}+C(x^3-3xy^2)-\frac{\sigma Qb^2x}{3EI}\\ &\quad+\frac{4\sigma b^3Q}{\pi^3EI}\sum_{n=1}^{n=\infty}\frac{(-1)^{n-1}}{n^3}\frac{\sinh\frac{n\pi x}{b}}{\cosh\frac{n\pi a}{b}}\cos\frac{n\pi y}{b}\\ \lambda_2 &= -\frac{\sigma Qy^3}{3EI}.\end{aligned}\right\}\quad(29'')$$

These equations are obtained from (29) by omitting terms containing B, S, τ and e and by adding the two last terms in the expression for λ_1. The latter terms satisfy by themselves the equation $\nabla^2\lambda_1=0$ and, as shown below, satisfy also the boundary conditions.

The equations of the boundaries are $x=\pm a$, $y=\pm b$ and the boundary conditions are that $q=0$ when $x=\pm a$ and $s=0$ when $y=\pm b$.

From(29′) we find by differentiation: $\frac{\partial\lambda_1}{\partial x}$, $\frac{\partial\lambda_1}{\partial y}$, $\frac{\partial\lambda_2}{\partial x}$ and $\frac{\partial\lambda_2}{\partial y}$

and substituting in:

$$q=G\left[\frac{\partial\lambda_1}{\partial x}-\frac{\partial\lambda_2}{\partial y}-b_1\right]\quad\text{and}\quad s=G\left[\frac{\partial\lambda_1}{\partial y}+\frac{\partial\lambda_2}{\partial x}\right]\quad(25')$$

where the symbol b_1 is used for b and terms in e and c are omitted we get:

$$q=G\left[-\frac{Qy^2}{2IG}+3C(x^2-y^2)+\frac{\sigma Q}{EI}y^2-\frac{Q}{3}\frac{\sigma b^2}{EI}+\frac{4\sigma b^3Q}{\pi^3EI}\sum_{n=1}^{n=\infty}\frac{(-1)^{n-1}}{n^3}\frac{\cosh\frac{n\pi x}{b}}{\cosh\frac{n\pi a}{b}}\cos\frac{n\pi y}{b}-b_1\right]$$

$$s=G\left[-\frac{Qxy}{IG}-6Cxy-\frac{4\sigma b^2Q}{\pi^2EI}\sum_{n=1}^{n=\infty}\frac{(-1)^{n-1}}{n^2}\frac{\sinh\frac{n\pi x}{b}}{\cosh\frac{n\pi a}{b}}\sin\frac{n\pi y}{b}\right]$$

Substitute now $x=\pm a$ in the expression for q and $y=\pm b$ in the expression for s and equate each of them to zero, noting that

$$\frac{b^2}{3}-\frac{4b^2}{\pi^2}\sum_{n=1}^{n=\infty}\frac{(-1)^{n-1}}{n^2}\cos\frac{n\pi y}{b}$$

is the expansion of y^2 in a Fourier's series, we find:

$$-\frac{Qy^2}{2IG}+3C(a^2-y^2)+\frac{\sigma Q}{EI}y^2-\frac{\sigma Q}{EI}y^2-b_1=0$$

$$\mp\frac{Qxb}{IG}\mp 6Cxb=0$$

from which, equating constant terms and terms in x or y^2:

$$b_1=-\frac{Qa^2}{2IG} \qquad C=-\frac{Q}{6IG}$$

Substituting in the expressions for q and s, and putting $I=\frac{4}{3}a^3b$ we get:

$$q=\frac{3q_0}{2(1+\sigma)a^2}\left\{\begin{array}{l}(1+\sigma)(a^2-x^2)-\sigma\left(\frac{b^2}{3}-y^2\right)\\ \qquad+\frac{4\sigma b^2}{\pi^2}\sum_{n=1}^{n=\infty}\frac{(-1)^{n-1}}{n^2}\frac{\cosh\frac{n\pi x}{b}}{\cosh\frac{n\pi a}{b}}\cos\frac{n\pi y}{b}\end{array}\right\} \quad (53)$$

$$s=-\frac{3q_0}{2(1+\sigma)a^2}\left\{\frac{4\sigma b^2}{\pi^2}\sum_{n=1}^{n=\infty}\frac{(-1)^{n-1}}{n^2}\frac{\sinh\frac{n\pi x}{b}}{\cosh\frac{n\pi a}{b}}\sin\frac{n\pi y}{b}\right\}$$

VI. *Beam of Rectangular Section. Torsion.*

In this case the terms in λ_1 given in (29) all disappear and we assume $\lambda_1=0$. In (28) the terms in S and Q disappear and this equation takes the form:

$$\frac{\partial^2\lambda_2}{\partial x^2}+\frac{\partial^2\lambda_2}{\partial y^2}=2(\tau+e). \quad (28')$$

In the expression for λ_2 it is now necessary to add a term in y^2 and a term consisting of a transcendental series, which by itself satisfies the equation $\nabla^2\lambda_2=0$. The term in y^2 is found to be necessary in order that q shall satisfy the boundary condition. We write:

$$\lambda_2=(\tau+e)(2x^2-y^2)-\left(\frac{4}{\pi}\right)^3\frac{\theta a^2}{2}\sum_{n=1}^{n=\infty}\frac{(-1)^{n-1}}{(2n-1)^3}\frac{\cosh\dfrac{(2n-1)\pi y}{2a}}{\cosh\dfrac{(2n-1)\pi b}{2a}}\times$$

$$\cos\frac{(2n-1)\pi x}{2a}\ldots \qquad (29')$$

where θ is an unknown constant.

The function λ_2 in this case has several interesting properties the discussion of which, however, lies outside the scope of this paper.[7]

From (25)

$$q=G\left[-\frac{\partial\lambda_2}{\partial y}-ey\right] \qquad \text{and } s=G\left[\frac{\partial\lambda_2}{\partial x}-ex\right].$$

Substituting for the derivatives of λ_2 as found from (29′) and making $x=\pm a$ in the expression for q, we have $q=0$ and since the cosine in the transcendental series vanishes, we obtain the following equation for the constants:

$$2(\tau+e)+e=0 \quad \therefore \quad e=-\tfrac{2}{3}\tau.$$

Making $y=\pm b$ we have $s=0$, whence:

$$4(\tau+e)x+\left(\frac{4}{\pi}\right)^3\theta a\sum_{n=1}^{n=\infty}\frac{(-1)^{n-1}}{(2x-1)^2}\sin\frac{(2n-1)\pi x}{2a}-ex=0$$

but the middle term is the expansion in a Fourier's series of x multiplied by 2θ

$$\therefore \quad 4(\tau+e)+2\theta-e=0$$

and substituting for e, we find $\theta=-\tau$.

[7] A. Ostenfeld, Teknisk Elasticitetslaere, Copenhagen, 1924.

Hence we can write the expressions for the stresses as follows:

$$\left.\begin{aligned} q &= -\tau aG\left(\frac{4}{\pi}\right)^2 \sum_{n=1}^{n=\infty} \frac{(-1)^{n-1}}{(2n-1)^2} \frac{\sinh \dfrac{(2n-1)\pi y}{2a}}{\cosh \dfrac{(2n-1)\pi b}{2a}} \cos \frac{(2n-1)\pi x}{2a} \\ s &= 2\tau Gx - \tau aG\left(\frac{4}{\pi}\right)^2 \sum_{n=1}^{n=\infty} \frac{(-1)^{n-1}}{(2n-1)^2} \frac{\cosh \dfrac{(2n-1)\pi y}{2a}}{\cosh \dfrac{(2n-1)\pi b}{2a}} \times \\ &\quad \sin \frac{(2n-1)\pi x}{2a}. \end{aligned}\right\} \quad (54)$$

Prandtl has shown[8] that the distribution of stresses given by these formulas can be exhibited to the eye by a homogeneous membrane under uniform pressure, as for instance a soap film, stretched over an opening of the same form as the section of the beam.

The constant τ is the torsional angle which is determined from (11). We shall here merely give the expression for T from which τ can be found when the torsional moment is given.

$$T = -\tau ba^3 G\left[\frac{16}{3} - \left(\frac{4}{\pi}\right)^5 \frac{a}{b} \sum_{n=1}^{n=\infty} \frac{1}{(2n-1)^5} \tanh \frac{(2n-1)\pi b}{2a}\right]. \quad (55)$$

An equivalent expression for T is obtained by simply interchanging a and b and should be used when $a > b$ as it is then more rapidly convergent than (55).

By means of tables for hyberbolic sines and cosines it is easy to find approximate values of q and s and hence of the resultant shear stress at any point to a reasonable degree of accuracy. The maximum stress is found at the middle of the longest side of the rectangular section. For a beam of square section the maximum stress is:

$$\tau_{\max.} = .600\frac{T}{a^3}. \quad (56)$$

[8] L. Prandtl, Phys. Zeitschrift, Bd. 4, 1903. Aug. Föppl and Ludwig Föppl "Drang und Zwang," 1920.

DETERMINANTAL EXPRESSION OF MULTIPLE CROSS AND DOT PRODUCTS OF POLYADICS IN THREE DIMENSIONS

By LEPINE HALL RICE

1. The double cross and double dot products of dyadics introduced by Gibbs have since been generalized and extensive study has been made of the p-tuple cross and dot products of p-adics. The p-tuple cross, or dot, product of two p-ads is the indeterminate, or scalar, product of cross, or dot, products of corresponding vectors; *e.g.*,

$$\mathbf{abc}\times_3\mathbf{fgh}=\mathbf{a}\times\mathbf{f}\quad\mathbf{b}\times\mathbf{g}\quad\mathbf{c}\times\mathbf{h},\text{ a triad;}$$

$$\mathbf{abc}\cdot_3\mathbf{fgh}=\mathbf{a}\cdot\mathbf{f}\quad\mathbf{b}\cdot\mathbf{g}\quad\mathbf{c}\cdot\mathbf{h},\text{ a scalar.}$$

Gibbs considered the scalar triple product

$$\Phi\overset{\times}{\times}\Psi:\Omega$$

of three dyadics and showed that this product is entirely independent of the order in which the factors are arranged and of the position of the dots and crosses. It is now known that if the three dyadics be written in nonion form, *viz.*,

$$\Phi=a_{11}\mathbf{ii}+a_{12}\mathbf{ij}+a_{13}\mathbf{ik}+a_{21}\mathbf{ji}+\ \ldots\ +a_{33}\mathbf{kk},$$

$$\Psi=b_{11}\mathbf{ii}+\ \ldots\ +b_{33}\mathbf{kk},\quad\Omega=c_{11}\mathbf{ii}+\ \ldots\ +c_{33}\mathbf{kk},$$

then the scalar triple product may be expressed as a cubic determinant whose strata (layers of the nonsignant[1] direction) are the 2-way matrices of the a's, the b's, and the c's, *i.e.*,

$$\Phi\overset{\times}{\times}\Psi:\Omega=\left[\begin{matrix}+\\ \pm\\ \pm\end{matrix}\right|\begin{matrix}a_{11}&a_{12}&a_{13}\\ a_{21}&a_{22}&a_{23}\\ a_{31}&a_{32}&a_{33}\end{matrix}\left|\begin{matrix}b_{11}&b_{12}&b_{13}\\ b_{21}&b_{22}&b_{23}\\ b_{31}&b_{32}&b_{33}\end{matrix}\right|\begin{matrix}c_{11}&c_{12}&c_{13}\\ c_{21}&c_{22}&c_{23}\\ c_{31}&c_{32}&c_{33}\end{matrix}\Bigg|\ .$$

This mode of expression again makes evident the above-stated independence of order.

1 The meaning of "signant" and "nonsignant" is given in the author's paper on P-way Determinants, American Journal of Mathematics, Vol. xl, (1918), p. 242, at p. 246. There the symbols "$\frown$" and "$\smile$" were used to mean respectively "signant" and "nonsignant"; but in later papers the author has used instead "$\pm$" and "$+$" either prefixed to the first bar (as here) or written over the signant and nonsignant indices. It will be observed, that in the present paper the generalized definition of a determinant in the cited paper is not involved, but the terminology there introduced is here useful.

2. We are naturally led to inquire in what determinantal forms the p-tuple cross-product of two p-adics and the p-tuple scalar triple product of three p-adics may be expressed.

Let

$$\mathbf{i}=\mathbf{e}_1, \quad \mathbf{j}=\mathbf{e}_2, \quad \mathbf{k}=\mathbf{e}_3;$$

$$\mathbf{A}=\Sigma a_{\alpha_1\alpha_2\ldots\alpha_p}\mathbf{e}_{\alpha_1}\mathbf{e}_{\alpha_2}\ldots\mathbf{e}_{\alpha_p}, \quad \alpha_i=1, 2, 3$$

$$\mathbf{B}=\Sigma b_{\beta_1\beta_2\ldots\beta_p}\mathbf{e}_{\beta_1}\mathbf{e}_{\beta_2}\ldots\mathbf{e}_{\beta_p}, \quad \beta_i=1, 2, 3$$

$$\mathbf{C}=\Sigma c_{\gamma_1\gamma_2\ldots\gamma_p}\mathbf{e}_{\gamma_1}\mathbf{e}_{\gamma_2}\ldots\mathbf{e}_{\gamma_p}, \quad \gamma_i=1, 2, 3\,.$$

Starting to form the p-tuple cross product $\mathbf{A}\times_p\mathbf{B}$, or $\mathbf{Q}$ say, we have, on applying the distributive law,

$$\mathbf{Q}\equiv\mathbf{A}\times_p\mathbf{B}=\Sigma a_{\alpha_1\ldots\alpha_p}b_{\beta_1\ldots\beta_p}(\mathbf{e}_{\alpha_1}\times\mathbf{e}_{\beta_1})\ldots(\mathbf{e}_{\alpha_p}\times\mathbf{e}_{\beta_p}).$$

Now it is easily verified that

$$\mathbf{e}_\lambda\times\mathbf{e}_\mu=(\pm_{\lambda\mu\nu})\mathbf{e}_\nu,$$

where $\lambda\neq\mu$, and $\lambda\mu\nu$ is a permutation of 1 2 3, and $(\pm_{\lambda\mu\nu})$ is Cayley's symbol having the value "+" or "−" according as the permutation is even or odd. Hence

$$\mathbf{Q}=\Sigma(\pm_{\alpha_1\beta_1\gamma_1})\ldots(\pm_{\alpha_p\beta_p\gamma_p})a_{\alpha_1\ldots\alpha_p}b_{\beta_1\ldots\beta_p}\mathbf{e}_{\gamma_1}\ldots\mathbf{e}_{\gamma_p}.$$

where, for every value of i from 1 to p, $\alpha_i\beta_i\gamma_i$ is a permutation of 1 2 3.

It will be seen that $\mathbf{Q}$ is a determinant of the $(p+1)$−way matrix of the third order whose (p-way) layers of the first direction are the matrices

$$\|a_{\alpha_1\ldots\alpha_p}\|, \quad \|b_{\beta_1\ldots\beta_p}\|, \quad \|(\mathbf{e}_{\gamma_1}\ldots\mathbf{e}_{\gamma_p})\|.$$

If p is odd it is evident that $\mathbf{Q}$ must have all directions signant, while if p is even $\mathbf{Q}$ must have all directions except the first signant. Writing $\pm^p$ to mean + (nonsignant) when p is even and $\pm$ (signant) when p is odd, the signancy just specified may be indicated by the symbol $(\pm^p\pm\ldots\pm)$ and both cases will be included in the theorem

$$\mathbf{Q}\equiv\mathbf{A}\times_p\mathbf{B}=(\pm^p\pm\ldots\pm)\,|\,a_{\alpha_1\ldots\alpha_p}\,|\,b_{\beta_1\ldots\beta_p}\,|(\mathbf{e}_{\gamma_1}\ldots\mathbf{e}_{\gamma_p})|^{(3)}_{p+1} \quad (1)$$

3. As particular cases, first if $p=1$ we have the well-known expression

$$\mathbf{a}\times\mathbf{b}=\begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ \mathbf{i} & \mathbf{j} & \mathbf{k} \end{vmatrix};$$

while if $p=2$ we have

$$\Phi \overset{\times}{\times} \Psi = \begin{bmatrix} + \\ \pm \\ \pm \end{bmatrix} \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} \begin{vmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{vmatrix} \begin{vmatrix} \mathbf{ii} & \mathbf{ij} & \mathbf{ik} \\ \mathbf{ji} & \mathbf{jj} & \mathbf{jk} \\ \mathbf{ki} & \mathbf{kj} & \mathbf{kk} \end{vmatrix}.$$

4. Completing the formation of the scalar triple product by operating on $\mathbf{A}\times_p\mathbf{B}$ with $\cdot_p\mathbf{C}$, it is obvious that we shall obtain as a second theorem

$$\mathbf{A}\times_p\mathbf{B}\cdot_p\mathbf{C}=(\pm^p\pm\ \cdots\pm)\,|a_{\alpha_1\ldots\alpha_p}|\,b_{\beta_1\ldots\beta_p}|c_{\gamma_1\ldots\gamma_p}|^{(3)}_{p+1}, \qquad (2)$$

that is, the scalar triple product of three p-adics is a determinant of the $(p+1)$-way matrix of the third order whose layers of the first direction are the p-way matrices of the factors, the signancy being as above specified.

5. It is readily verified that the scalar triple product is in all cases independent of the order of the crosses and dots. If p is even, the product is entirely independent of the order of the factors, while if p is odd the product is unchanged by a cyclic interchange of the factors but changes sign upon a reversal of the cyclic order.

If p is odd and a linear relation exists among the factors, it follows from the properties of determinants that the scalar triple product vanishes. In particular, it vanishes if two factors are alike.

NOTE ON THE SCATTERING OF X-RAYS

By P. Debye[1]

I. The phenomena observed by v. Laue and Bragg give us evidence of two fundamental facts:

1. They show that X-rays, scattered by an atom, must have, at least partly, phase relations with the primary rays,[2] producing the well-known interference effects.

2. They show that the intensity of scattering depends on the inner properties of the atoms inducing scattering.

There can be no question that not only in crystals, but also in the molecules of substances in the liquid or gaseous state, the atoms occupy definite places. It is true that in general the distances of the atoms will not be quite constant, but certain types of inner vibrations will exist. The values for the specific heat at constant volume for more complicated molecules give evidence of this heat motion. It seems, however, reasonable to admit that these inner vibrations will only have a disturbing effect of a second order of magnitude, as has been found in the case of the temperature effect in crystal reflection.

Moreover, there seems to be no reason why the atoms in a molecule should act in another manner than the atoms in a crystal. This is obvious if the crystal is considered as one large molecule, the picture of the lattice constituting its chemical structure formula, and if it is borne in mind that the interatomic forces in the crystal are of the same kind as the interatomic forces which help to build up the molecule.

Therefore it should be possible to detect the interference effects corresponding to the geometrical arrangement of the atoms in the molecule.

In the theory of v. Laue it is assumed that each atom acts as

[1] Professor of Physics at the Eidgenössische Technische Hochschule, Zürich, Lecturer at the Department of Physics, Massachusetts Institute of Technology, Winter Term, 1924–1925.

[2] Experiments of G. E. M. Jauncey and C. H. Eckart, Nature, 112, Sept., 1923; H. Kulenkampff, Zeits. f. Physik, 19, 1923; and Y. H. Woo, Proc. Nat. Acad., 10, 145, 1924, seem to show that the scattered radiation with shifted wave-length, detected by Compton, is not able to give interference effects of the kind necessary for the crystal X-ray reflection.

a resonator, scattering the primary radiation in all directions but with an amplitude and phase which may depend, in general, on the constitution of the atom and the wave-length of the primary rays. Laue himself does not make any attempt to find a connection between the atomic properties and their scattering coefficient, this coefficient being introduced as a quantity only phenomenologically defined. In the later work of C. G. Darwin, P. Debye and A. H. Compton, an attempt is made to give a theory of the properties of this coefficient, assuming with J. J. Thomson that each electron in the atom is the seat of the fundamental scattering effect, and the whole atomic effect is really the result of a summation extending over all the electrons included in the atom. The discussion of some early experimental work of Debye and Scherrer on the scattering by diamond, and the later, more finished work of W. L. Bragg on rock salt, is based on this theory.

It is not an easy matter to derive the atomic scattering coefficient from observations of the intensity of the reflected beam, the only thing really measured. It can only be done with the aid of a complicated and very elaborate theory consisting of a summation extending over the elementary wavelets issuing from the single atoms. This theory itself is not beyond question; very recently P. Ewald[3] has been led to a different final result, taking account of the interaction of the wavelets more decidedly than in the first theory of v. Laue.

At first glance it seems that we may avoid this difficulty if the experiments are not performed with substances in the crystalline state. If, for instance, the scattering of a liquid has been measured, one might possibly think that the properties of the scattering function, *i.e.*, the scattered intensity plotted as a function of the angle between the secondary and primary X-ray, only depend on the dimensions and the form of the atomic frame constituting the molecule. From this point of view P. Debye and P. Scherrer in 1916 discussed the scattering of benzene.[4] In the meantime it became evident, however, that a very large number of different liquids yield diffraction-patterns which show only slight differences. Therefore, at the meeting of the German Physical

[3] P. Ewald, Phys. Zeitschr., 25, 29, 1925.
[4] Gött. Nachr., 16, 1916.

Society in 1920 at Jena, Debye stated that the principal maximum of the scattering function (*i.e.*, the first interference ring surrounding the primary ray) must be due to interferences between the different molecules of the liquid. Experiments carried out by W. H. Keesom[5] which led him to the same statement — especially the fact that he was able to photograph the ring even with a monatomic gas in the liquid state — are conclusive.

The theory of this effect seems to be as difficult as the theory of the correction for the dimensions of the molecules in the equation of state. Nevertheless, I believe that a first approximation, which can readily be obtained for the limiting case that the total volume of the molecules is small compared with the volume occupied by the gas, may be of some interest. It will be shown that, even if the molecules are comparable with hard spheres and do not interact in any other way than to prevent each other from entering into the domain defined by this sphere, this fact alone is sufficient to cause a scattering function, exhibiting a maximum. This maximum occurs at an angle defined by the quotient of the wave-length and the diameter of the sphere substituting the molecule.

In this way it seems proved, both experimentally and theoretically, that it will only be possible, even in the case of liquids, to arrive at the scattering function characteristic of the molecule and its atomic frame, if we succeed in freeing the primary experimental result of the undesired intermolecular interferences. Further, it seem improbable that it will be possible to perform theoretical calculations which will give a reliable formula for this process in the case of dense liquids. Therefore the only possible way to find the interferences due to the interaction of the atoms constituting the molecule, seems to be the performance of scattering experiments with gases of different densities. If such experiments are done, then, with the help of theoretical considerations of the kind given below, it should be possible to eliminate the effect of molecular interaction, and to arrive at the direct measurement of the atomic distances in the molecule.

[5] W. H. Keesom and J. de Smedt, Journ. de Phys. Ser. VI, 4, 144, 1923.

II. Suppose a beam of primary X-rays, traveling in such a direction that the cosines of the angles with the three axes of a rectangular system of coördinates are α_0, β_0, γ_0 and define this

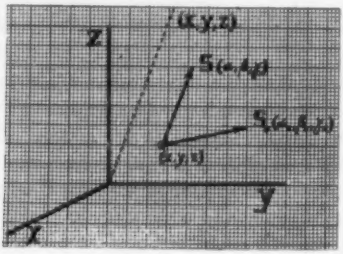


Fig. 1

direction by a vector S_0 of unit length and components α_0, β_0, γ_0. Let this ray strike a particle at a point x, y, z and let us assume that this particle will give rise to a scattered radiation in all directions with amplitude and phase defined in the usual way by a scattering factor Ψ. If the amplitude of the primary ray equals 1, the electric force at any point x, y, z, and any time t is equal to

$$e^{i\omega t}e^{-ik(\alpha_0 x+\beta_0 y+\gamma_0 z)}$$

if ω is the frequency and

$$k=\frac{\omega}{c}=\frac{2\pi}{\lambda}$$

is defined by the frequency ω and the velocity of propagation c or by the wave length λ. The particle at the point x, y, z will therefore emit a secondary wave with amplitude and phase given by the factor

$$\Psi e^{-ik(\alpha_0 x+\beta_0 y+\gamma_0 z)}.$$

The electric force of this scattered wave in a point X, Y, Z at a relatively large distance l, from the particle will be equal to

$$\Psi e^{-ik(\alpha_0 x+\beta_0 y+\gamma_0 z)}\ \frac{e^{-ikl}}{l}.$$

Suppose now the point X, Y, Z at so large a distance from the scattering medium that, introducing the distance R of this point

from the origin, and the cosines α, β, γ of the radius vector R with the axes, we may use the approximation

$$l = R - (\alpha x + \beta y + \gamma z).$$

The amplitude of the scattered wave will then be equal to

$$\Psi \frac{e^{-ikR}}{R} e^{ik[(\alpha-\alpha_0)x + (\beta-\beta_0)y + (\gamma-\gamma_0)z]}.$$

Introducing a vector S of unit length and the components α, β, γ, and the vector r with the components x, y, z, defining the place of the particle, this expression can be written in the form

$$\Psi \frac{e^{-ikR}}{R} e^{ik(S-S_0).r}$$

if $(S-S_0).r$ is the scalar product of the two vectors $S-S_0$ and r.

Suppose now, that the scattering medium consists of a large number of particles $1 \ldots n \ldots N$, at different places, characterized at a certain moment by the vectors $r_1 \ldots r_n \ldots r_N$. At this moment the amplitude of the scattered radiation at the point X, Y, Z will then be equal to

$$\Psi \frac{e^{-ikR}}{R} \sum_n e^{ik(S-S_0).r_n}. \tag{1}$$

But the intensity corresponding to this amplitude cannot really be observed if the particles of the medium have a heat motion, as for instance, the molecules of a gas or a fluid. The only thing which can be measured is the mean value of this intensity during an appreciable time, very large in comparison with the time necessary for the positions of the particles to become thoroughly interchanged.

According to (1) we will find the intensity I at the moment under consideration by multiplying this expression by the conjugate value, finding

$$I = \frac{\Psi^2}{R^2} \sum_m \sum_n e^{ik(S-S_0).(r_n - r_m)}, \tag{2}$$

if we treat Ψ as a real value. The intensity to be measured I_m, will be given by the mean value of this expression, an operation which will be indicated by the letter M, so that

$$I_m = \frac{\Psi^2}{R^2} \sum_m \sum_n M[e^{ik(S-S_0).(r_n - r_m)}]. \tag{3}$$

III. If the number of particles of the scattering gas is equal to N, the sum in (3) consists of N^2 terms. The N terms of this sum, in which $n=m$, each have the value 1; the $N(N-1)$ remaining terms depend on the relative positions of two particles n and m. We now have to calculate the mean value

$$M[e^{ik(S-S_0)\cdot(r_n-r_m)}]=M_{mn}$$

giving the particles all possible places in the gas. We introduce the assumption that the gas considered is not very dense, so that, dealing with two particles m and n, we need not take into account the other particles. Moreover, it will not be necessary to make any difference between the real volume of the gas and the free volume in which the molecules can move. Calling the whole volume occupied V and an element of this volume dV, the probability that the center of the particle n will be in dV_n and the center of the particle m in dV_m is equal to

$$\frac{dV_n}{V}\ \frac{dV_m}{V}\ .$$

Multiplying the expression $e^{ik(S-S_0)\cdot(r_n-r_m)}$ by $\frac{dV_n}{V}\ \frac{dV_m}{V}$ and integrating over all the possible situations will yield the desired mean value M_{mn}.

If we fix the molecule n and allow the molecule m to reach every point not forbidden by our fundamental assumption that each molecule will act as a hard sphere with a certain radius a, the integration according to dV_m must be performed over the whole volume, with the exception of a sphere of the radius $2a$ surrounding the molecule n. This integration achieved, we will have to move dV_n to all the places included in V. Instead of performing the integration in this way, we may also integrate first by taking into account all the elements of volume dV_n and dV_m, of V without making any exception, corresponding to the molecular volume. We will then have to subtract the result issuing from the integration with respect to dV_m, now moving dV_m only in the interior of the sphere, with radius $2a$, followed by an integration with respect to dV_n over all the elements contained in V. This is the method we will follow.

The result therefore assumes the form:

$$M_{mn} = K'_{mn} - K''_{mn} \tag{4}$$

with

$$K'_{mn} = \int_{\text{Vol.}} \int_{\text{Vol.}} \frac{dV_n}{V} \frac{dV_m}{V} e^{ik(S-S_0)\cdot(r_n - r_m)} \tag{5}$$

and

$$K''_{mn} = \int_{\text{Vol.}} \int_{\text{Sphere}} \frac{dV_n}{V} \frac{dV_m}{V} e^{ik(S-S_0)\cdot(r_n - r_m)} \tag{5'}$$

the suffixes vol. and sphere at the integrals referring to the domains over which the integration has to be performed.

IV. Putting $S-S_0=s$, $r_n-r_m=r$ and introducing the angle θ between the vectors $S-S_0$ and r_n-r_m, the inner integral in (5') assumes the form

$$\frac{2\pi}{V} \int_{r=0}^{r=2a} r^2 dr \int_{\theta=0}^{\theta=2\pi} e^{iksr\cos\theta} \sin\theta \, d\theta.$$

The result is

$$\frac{4\pi}{V} (2a)^3 \frac{\sin 2ksa - 2ksa \cos 2ksa}{(2ksa)^3},$$

which we will write in the form

$$\frac{1}{V} \frac{4\pi}{3} (2a)^3 \, \phi \, (2ksa),$$

introducing the function

$$\phi(u) = \frac{3}{u^3} (\sin u - u \cos u) \tag{6}$$

being equal to 1 for $u=0$.

The calculation of the outer integral in (5') means merely the multiplication with the factor 1. Therefore we arrive at

$$K''_{mn} = \frac{1}{V} \frac{4\pi}{3} (2a)^3 \, \phi \, (2ksa). \tag{7}$$

From a purely mathematical point of view the calculation of

K'_{mn} is very difficult, because the expression will depend upon the geometrical form of the volume V. From a physical point of view it is, however, evident that this geometrical form can practically be of no importance at all. To find the order of magnitude of the value designated by K'_{mn} we will calculate the integral for a volume of definite form, *i.e.*, a cube of side d.

Let the coördinates of r_m and r_n be called x_m, y_m, z_m and x_n, y_n, z_n, the coördinates of S and S_0 being α, β, γ and α_0, β_0, γ_0. If the corner of the cube coincides with the origin of the system of coördinates and its sides are parallel to the axes, then we find, according to (5)

$$K'_{mn} = \frac{1}{d^6} \int_0^d \cdots$$

$$\int_0^d e^{ik[(\alpha-\alpha_0)(x_n-x_m)+(\beta-\beta_0)(y_n-y_m)+(\gamma-\gamma_0)(z_n-z_m)]}\, dx_n dy_n dz_n dx_m dy_m dz_m.$$

Now we have at once

$$\int_0^d \int_0^d e^{ip(\xi_1-\xi_2)} d\xi_1 d\xi_2 = 2\frac{1-\cos pd}{p^2}$$

so that

$$K'_{mn} = 8\frac{1-\cos kd(\alpha-\alpha_0)}{k^2d^2}\ \frac{1-\cos kd(\beta-\beta_0)}{k^2d^2}\ \frac{1-\cos kd(\gamma-\gamma_0)}{k^2d^2} \quad (8)$$

It is obvious that K'_{mn} will be equal to 0 for very small values of the angle between the scattered and the primary ray because $kd = 2\pi\frac{d}{\lambda}$ is always an exceedingly large number. On the other hand, K''_{mn} depends on the value of ksa. The diameter of the molecule a being of the same order of magnitude as the wave length, this expression will have an appreciable value even for large angles.[6] We can therefore practically put

$$K'_{mn} = 0 \quad (8')$$

[6] As is readily seen the quotient K'_{mn}/K''_{mn} is of the order of magnitude volume of one molecule / total volume of the gas.

V. With the help of (7) and (8′), we can calculate

$$M_{mn} = -\frac{1}{V}\frac{4\pi}{3}(2a)^3\,\phi\,(2ksa).$$

M_{mn} not depending on the suffixes, and entering $N(N-1)$ times in the sum (3), expressing I_m, we find

$$I_m = \frac{\Psi^2}{R^2}\left[N - \frac{N(N-1)}{V}\frac{4\pi}{3}(2a)^3\phi(2ksa)\right].$$

The product

$$N\frac{4\pi}{3}(2a)^3 = \Omega$$

is the total volume of all the "spheres of action" of all the molecules, each sphere having a radius equal to the diameter of one molecule. Moreover, N being always very large, we may put $N-1=N$. The intensity of the scattering, therefore, is

$$I_m = N\frac{\Psi^2}{R^2}\left[1 - \frac{\Omega}{V}\phi\,(2ksa)\right]. \tag{9}$$

We put $|S-S_0|=s$; if now the angle between the secondary and the primary ray is called θ, then

$$s^2 = (S-S_0)^2 = S^2 + S_0^2 - S.S_0 = 2(1-\cos\theta) = 4\sin^2\frac{\theta}{2}.$$

The argument of our function ϕ, defined by (6) can, as a result, be written in the form

$$2ksa = 8\frac{\pi a}{\lambda}\sin\frac{\theta}{2}$$

remembering that $k = \frac{2\pi}{\lambda}$. According to (9), the intensity of scattering, therefore, depends on the angle θ, even for the case underlying our calculation, that the molecules only act as hard spheres and show no association at all to double molecules. This remark seems necessary because it has sometimes been claimed that the existence of the interference ring has to be connected with the existence of double molecules, no other cause seeming to be available for the understanding of interference effects.

In Fig. 2 the function

$$\phi(u)=\frac{3}{u^3}(\sin u-u\cos u)$$

is plotted as an ordinate against u. For $u=0$ it is found $\phi(u)=1$, with increasing argument the function ϕ becomes 0, performing oscillations of decreasing amplitude. If the scattering of each

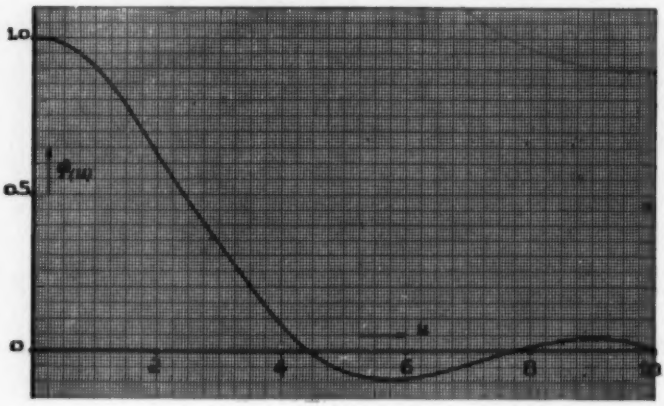


Fig. 2

particle were the same as the scattering of a single resonator, Ψ^2 would have to be proportional to $\frac{1+\cos^2\theta}{2}$ because of the fact that the primary radiation is unpolarized and the effect of scattering yields a polarization of the scattered radiation which is complete for an angle of $\theta=90°$. To visualize the effect of the interference here considered, a curve is drawn in Fig. 3, giving the function

$$\frac{1+\cos^2\theta}{2}\left[1-\tfrac{1}{2}\phi\left(12\,\pi\,\sin\frac{\theta}{2}\right)\right].$$

This corresponds to

$$\frac{2a}{\lambda}=3$$

meaning a diameter of $2a = 2.1 \times 10^{-8}$ cm., if a wave length of $\lambda = 0.7 \times 10^{-8}$ is used. Moreover, the factor $\frac{\Omega}{V}$ has been put equal to $\frac{\Omega}{V} = \frac{1}{2}$. Under these assumptions, a ring would be observed with a maximum of intensity at an angle of about $\theta = 16°$.

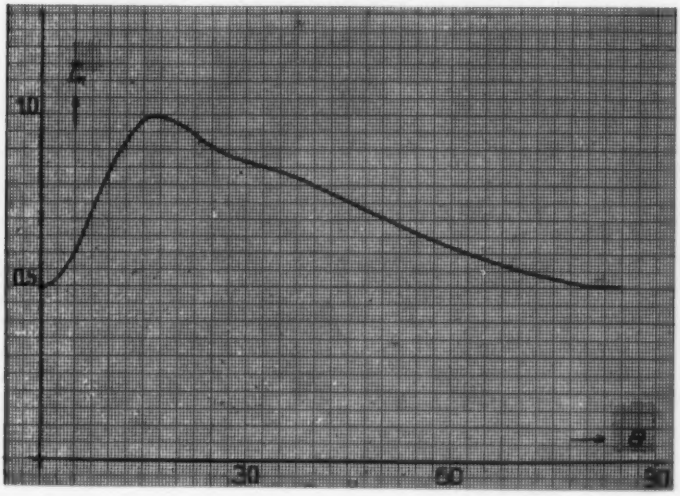


Fig. 3

In general, the molecule will not act as a single resonator, but it is to be expected that there will exist an interference effect corresponding to the shape of the atomic frame. The whole scattering curve must now show a superposition of the two kinds of interference effects which may be briefly called the "inner" and the "outer" effect. Now the outer effect will, as in (9'), always be proportional to $\frac{\Omega}{V}$. In making measurements on a gas under different pressures, it should therefore be possible to eliminate the outer effect which tends to vanish with decreasing density,

and to arrive at a measurement of the inner effect. Should such measurements prove to be practically possible, then we may, in the case of nitrogen for instance, expect as a result the direct measurement of the distance of the two N-atoms.

VI. It will, perhaps, be of some interest to find the scattering function also for this more complicated case. Consider spherical molecules of radius a, each containing two scattering centers acting as simple resonators at a mutual distance l. The line connecting these two centers is supposed to pass through the center of the sphere so that this center is at equal distance $\frac{l}{2}$ from the two resonators. The molecules themselves will be numerated in the same manner as before with the numbers $1 \ldots n \ldots N$. The resonators may be designated by dashes, in such a way that, for instance, r_n' means the radius-vector from the origin to one of the resonators of the particle n, and r_n'' means the radius-vector from the origin to the second resonator of the same molecule. The undashed letter r_n will be reserved for the vector drawn from the origin to the center of the sphere. Moreover, we shall put

$$r_n' = r_n + v_n' \qquad \text{and} \qquad r_n'' = r_n + v_n''$$

v_n' and v_n'' meaning the radius-vector drawn from the center of the sphere to the first or second resonator. According to our previous reasoning, (cf. (1)), the amplitude of the scattered radiation is now equal to

$$\frac{\Psi}{R} \sum_n \left(e^{ik(S-S_0)\cdot r_n'} + e^{ik(S-S_0)\cdot r_n''}\right)$$

and the mean intensity I_m can be calculated by the equation

$$I_m = \frac{\Psi^2}{R^2} \sum_m \sum_n M_{mn} \tag{10}$$

with

$$M_{mn} = M\left[\left\{e^{ik(S-S_0)\cdot r_n'} + e^{ik(S-S_0)\cdot r_n''}\right\}\left\{e^{ik(S-S_0)\cdot r_m'} + e^{ik(S-S_0)\cdot r_m''}\right\}\right]. \tag{10'}$$

Now the mean values of each of the four terms of this product must be calculated. Each of them will give another result according as to whether $n = m$ or $n \neq m$. Assuming that all the possible

directions of the line connecting the two resonators and therefore of the vectors v_n' and v_n'' are equally probable, this calculation however, offers no difficulty,[7] and leads to the following formula:

(*a*) $m=n$

$$M[e^{ik(S-S_0)\cdot(r_n'-r_m')}]=1$$

$$M[e^{ik(S-S_0)\cdot(r_n'-r_m'')}]=\frac{\sin ksl}{ksl}$$

$$M[e^{ik(S-S_0)\cdot(r_n''-r_m')}]=\frac{\sin ksl}{ksl}$$

$$M[e^{ik(S-S_0)\cdot(r_n''-r_m'')}]=1$$

(*b*) $m\neq n$

$$M[e^{ik(S-S_0)\cdot(r_n'-r_m')}]=-\frac{1}{N}\frac{\Omega}{V}\phi(2ksa)\left(\frac{\sin\frac{ksl}{2}}{\frac{ksl}{2}}\right)^2$$

and the same value for the three remaining terms. Remembering that in N terms of the sum (10) the condition $n=m$ is fulfilled and the number of the remaining terms in $N(N-1)$, we find first

$$n=m:\ M_{nm}=2\left[1+\frac{\sin ksl}{ksl}\right]$$

$$n\neq m:\ M_{nm}=-\frac{4}{N}\frac{\Omega}{V}\ \phi(2ksa)\left(\frac{\sin\frac{ksl}{2}}{\frac{ksl}{2}}\right)^2$$

[7] Calculating $M\,e^{ik(S-S_0)\cdot(r_n'-r_m')}$ we put

$$r_n'=r_n+v_n' \qquad , \qquad r_m'=r_m+v_m'$$

The mean value M to be found is then the result of the following integration

$$M=\int\int\frac{dV_n}{V}\frac{dV_m}{V}e^{ik(S-S_0)\cdot(r_n-r_m)}\ \int\int\frac{d\Omega_n}{4}\frac{d\Omega_m}{4}e^{ik(S-S_0)\cdot(v_n'-v_m')}$$

if $d\Omega_n$ and $d\Omega_m$ denote elements of the solid angles defining the directions of the vectors v_n' and v_m'. Remembering that the length of either of these vectors is $l/2$, and introducing the angle Θ between $S-S_0$ and v_n', one of the last integrals assumes the form

$$\tfrac{1}{2}\int_0^{\pi}e^{\frac{ikls}{2}\cos\Theta}\ \sin\Theta\,d\,\Theta=\frac{\sin\frac{ksl}{2}}{\frac{ksl}{2}}$$

The first integral has been calculated in the preceding part. The combination of these results yields the formula under (*b*) of the text.

and on introducing these values in (10):

$$I_m = 4N\frac{\Psi^2}{R^2}\left\{\tfrac{1}{2}\left[1+\frac{\sin ksl}{ksl}\right]-\frac{\Omega}{V}\left(\frac{\sin\frac{ksl}{2}}{\frac{ksl}{2}}\right)^2\phi(2ksa)\right\}. \quad (11)$$

It may be remembered that,

$$ksl = 4\pi\frac{l}{\lambda}\sin\frac{\theta}{2}$$

$$2ksa = 8\pi\frac{a}{\lambda}\sin\frac{\theta}{2}$$

and ϕ is the function defined by (6) and given in the diagram Fig. 2.

With decreasing density the influence of the second term in (11) vanishes and the scattering function finally assumes the limiting form:

$$I_m = 4N\frac{\Psi^2}{R^2}\ \frac{1}{2}\left[1+\frac{\sin ksl}{ksl}\right], \quad (11')$$

agreeing with a general formula previously calculated.[8]

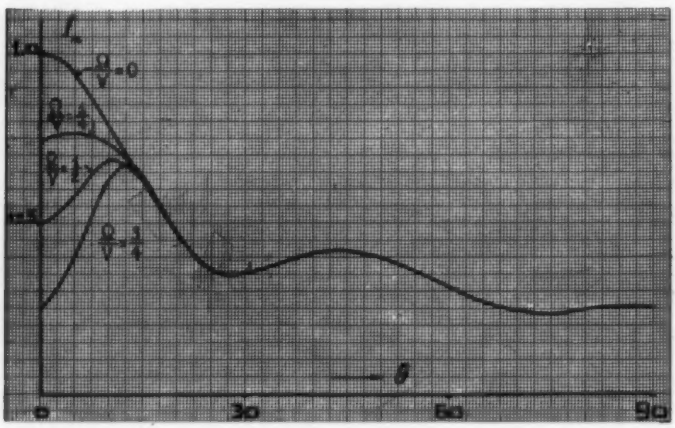


Fig. 4

[8] P. Debye, Ann. d. Phys. 46, 809, 1915.

To illustrate formula (11′) the intensity I_m has been calculated according to the assumptions $\frac{2a}{\lambda}=3$ and $l=a$, corresponding, for instance, at a given wave length of, say $\lambda=0.7\times10^{-8}$ cm. to a diameter of the sphere, substituting the molecule of 2.1×10^{-8} cm. and a distance of the resonators substituting the atoms of 1.05×10^{-8} cm. The scattering function moreover depends on the density, measured by the quotient Ω/V. Therefore four curves are given in Fig. 3, corresponding to the values $\frac{\Omega}{V}=0$, $\frac{\Omega}{V}=\frac{1}{4}$, $\frac{\Omega}{V}=\frac{1}{2}$ and $\frac{\Omega}{V}=\frac{3}{4}$. The diagram shows how the first maximum, occurring at about $\theta=12°$ and corresponding to intermolecular interference, disappears with decreasing density, leaving a second maximum at $\theta=45°$ undisturbed, because this maximum corresponds to the interatomic interference of the two atoms constituting the molecule.

March, 1925.

THE WEIERSTRASS APPROXIMATION THEOREM[1]

By PHILIP FRANKLIN

The theorem that a polynomial may be found which approximates a given continuous function uniformly to any degree of exactness is due to Weierstrass.[2] Of the numerous proofs which have since been given, probably the simplest and most elementary is that of Lebesgue.[3] While the corresponding theorem for functions of n variables is well known, the extension of the method of Lebesgue to this case, in all its simplicity, has so far not been given, and forms the object of the present note. Incidentally a method of approximating a continuous function of n variables by a polygonal function is developed which, it is believed, has other applications.

We shall begin by recalling the elementary proof for functions of one variable. Let the given function, $f(x)$, be continuous in the closed interval $a \leq x \leq b$. Since it is uniformly continuous in this interval, we can divide this interval into N equal parts such that the oscillation of the function in each of them is less than $\epsilon/2$. We now construct the broken line, or polygonal function, $G(x)$, which is linear inside each of these N intervals, and coincides in value with $f(x)$ at the points of subdivision. Evidently

$$|f(x)-G(x)| < \epsilon/2$$

throughout the interval.

If we define a ridge function as a function which is zero for $x \leq p$, linear for $x \geq p$, and continuous at the point $x = p$, we see that our polygonal function may be looked on as the sum of a constant and N ridge functions. For we may reduce it to zero by subtracting from it the constant $G(a)$ which makes it zero at a, then the ridge function coinciding with it in the first interval $a \leq x \leq a + \frac{b-a}{N}$, the ridge function coinciding with the remainder

[1] Presented to the American Mathematical Society, December 30, 1924.
[2] Berliner Sitzungsberichte, 1885.
[3] Lebesgue, Bulletin des Sciences Mathématiques, 1898, p. 278.

function in the second interval, and so on. But the ridge function which equals $A(x-p)$ for $x \geq p$, may be written:

$$R(x) = \frac{A}{2}\{x-p+|x-p|\}.$$

Furthermore

$$x-p = L\sqrt{1-\left[1-\left(\frac{x-p}{L}\right)^2\right]}$$

$$= L\left(1-\tfrac{1}{2}\left[1-\left(\frac{x-p}{L}\right)^2\right]-\tfrac{1}{8}\left[1-\left(\frac{x-p}{L}\right)^2\right]^2-\dots\right).$$

This is dominated by the convergent[4] series:

$$1+\tfrac{1}{2}+\tfrac{1}{8}+\dots+\frac{1.1.3\dots 2n-3}{n!\quad 2^n}+\dots$$

in the interval $p-L \leq x \leq p+L$. Consequently it is uniformly convergent there, and by taking a sufficient number of terms in the series we may obtain a polynomial approximating $|x-p|$, and hence one approximating the ridge function, as closely as we please in this interval. In particular, by taking $L > b-a$, we may find a polynominal approximating each of the ridge functions used to build up $G(x)$ to within $\epsilon/2N$ in an interval including the one $a \leq x \leq b$. Consequently, the sum of these polynomials will yield a polynomial $P(x)$ such that:

$$|G(x)-P(x)| < \epsilon/2$$

throughout this interval. By comparing this with the previous inequality, we see that

$$|f(x)-P(x)| < \epsilon, \qquad a \leq x \leq b,$$

and consequently $P(x)$ is the polynomial whose existence we sought to prove.

Consider next a function of two variables, $f(x, y)$, continuous in some closed, finite region. We enclose this region in a square,

$$a \leq x \leq b, \qquad c \leq y \leq d.$$

Since the function is uniformly continuous in its region of definition, we can divide this square into N^2 equal squares, such that the oscillation in each of them is less than $\epsilon/6$. We then divide

[4] Cf. Fine, College Algebra, p. 528.

each of these squares into two triangles by drawing the diagonal parallel to the line $x=y$. The polygonal function $G(x, y)$ will be built up on these triangles, and will first be defined at the vertices. At any vertex included in the region of definition of $f(x, y)$, we take $G(x, y)=f(x, y)$. At a vertex where $f(x, y)$ is not defined, but which abuts on one or more triangles containing a point where $f(x, y)$ does exist, we take for $G(x, y)$ any value assumed by $f(x, y)$ in an adjacent triangle. At the remaining vertices we take $G(x, y)=0$. We complete the definition of $G(x, y)$ by taking it inside any triangle as a linear function agreeing with the three values already assigned to the vertices. It is obviously continuous. Since the oscillation of $f(x, y)$ in any square, and therefore in any triangle is less than $\epsilon/6$, the function obtained by combining with $f(x, y)$ the values given to $G(x, y)$ at vertices where $f(x, y)$ is not defined, but which abut on triangles containing points where $f(x, y)$ does exist has its oscillation less than $\epsilon/2$. Consequently we have

$$|f(x, y)-G(x, y)|<\epsilon/2$$

in the entire region of definition of $f(x, y)$.

If we define a ridge function of two variables as a function which is zero for $x\leq p$ or $y\leq q$, is linear in each of the regions $x-p\leq y-q\leq 0$ and $y-q\leq x-p\leq 0$, and is continuous at all points, we see that our polygonal function $G(x, y)$ may be looked on as the sum of a constant, N ridge functions of x, N ridge functions of y, and N^2 ridge functions of x and y. For, we may reduce it to zero as follows. First subtract off a constant equal to $G(a, c)$ and a series of N ridge functions of x which reduces it to zero along the line $y=c$. Then subtract off N ridge functions of y which reduce the remainder function to zero along the line $x=a$. These subtractions clearly do not disturb its polygonal character. Then we subtract off the ridge function of x and y which coincides with it in the first square $a\leq x\leq a+\frac{b-a}{N}$, $c\leq y\leq c+\frac{b-a}{N}$. As the remainder function is now zero along the lines $x=a+\frac{b-a}{N}$, $y=c$, we may subtract off another ridge function of x and y which reduces it to zero in a second square, without dis-

turbing it in the first one. We keep on in this way until all the squares abutting on the line $y=c$ are traced, then we proceed to those abutting on the line $y=c+\frac{b-a}{N}$, and so on. This proves our contention about the composition of the polygonal function $G(x, y)$.

The approximation of ridge functions of a single variable by polynomials has already been given. For those involving two variables, we first notice that a ridge function with vertex at p, q may be written:

$$R(x, y)=\frac{A}{4}\{y-q+x-p-|y-q-x+p| \\ +|y-q+x-p-|y-q-x+p||\}$$

since this is zero for $x\leq p$ or $y\leq q$, is linear in each of the regions $x-p\leq y-q\leq 0$ and $y-q\leq x-p\leq 0$ being equal to $A(y-q)$ and $A(x-p)$ respectively, and is accordingly continuous at all points. To represent it by an approximating polynomial, we first observe that by the method used above for the absolute value of $x-p$, the absolute value of any function of x and y may be approximated as closely as desired by a polynomial in this function, in any region for the fundamental variables x and y in which this function remains bounded. Since this polynomial is uniformly continuous, we see that if the function itself may be approximated to any degree by a polynomial in x and y, its absolute value may be similarly approximated. As the expression for $R(x, y)$ given above is, to within a constant factor, obtained by adding together

$$y-q+x-p-|y-q-x+p|$$

which can be approximated by a polynomial by the previous argument, and its absolute value, which can be similarly approximated by the argument just given, we see that $R(x, y)$ may be approximated to any degree of exactness by a polynomial in x and y.

Since $G(x, y)$ was made up of a constant, $2N$ ridge functions of one variable, and N^2 ridge functions of two variables, if we obtain polynomials approximating each of these functions to within

$\epsilon/2(2N+N^2)$, their sum will be a polynomial $P(x, y)$ such that

$$|G(x, y)-P(x, y)|<\epsilon/2$$

throughout the entire fundamental square. Since this square includes the region of definition of $f(x, y)$, by comparing this with the earlier inequality, we see that in this region

$$|f(x, y)-P(x, y)|<\epsilon.$$

Thus $P(x, y)$ is the polynomial whose existence the theorem asserts.

The extension to functions of three and n variables is now obvious. The polygonal functions are here linear in each of the six ($n!$) triangular pyramids obtained from a cube (n-dimensional cube) by drawing in one diagonal which joins two opposite vertices, using one of the vertices as the common vertex of the pyramids, and the faces adjacent to the other vertex, subdivided into triangles by the projections of the diagonal (n-2 projections in the n-case) as the bases. All the diagonals drawn in are parallel, say to the line $x_1=x_2=\ldots x_n$. The ridge functions of $3(n)$ variables introduced at each stage are zero in the three (n) faces of the cube adjacent to one vertex, and in the $7(2^n-1)$ quadrants not containing the cube. They are linear in the three (n) square pyramids formed by combining the triangular ones previously mentioned, and continuous for all values. To represent them by absolute value signs, we have, putting

$$f_2=\tfrac{1}{4}\{x_2+x_1-|x_2-x_1|+|x_2+x_1-|x_2-x_1||\}$$

$$f_3=\tfrac{1}{4}\{f_2+x_3-|f_2-x_3|+|f_2+x_3-|f_2-x_3||\}$$

and, in general,

$$f_n=\tfrac{1}{4}\{f_{n-1}+x_n-|f_{n-1}-x_n|+|f_{n-1}+x_n-|f_{n-1}-x_n||\}$$

the ridge function of n variables $x_1 \ldots x_n$ as a constant multiple of f_n.

The polygonal functions $G(x_1 \ldots x_n)$ is built up of a constant and ridge functions of all orders up to the nth. The transition from the expressions in terms of absolute values to approximating polynomials is an obvious extension of the argument given above for the two variable case.

THE SOLUTION OF A DIFFERENCE EQUATION BY TRIGONOMETRICAL INTEGRALS

By NORBERT WIENER

Introduction

1. In a paper soon to appear,[1] the author has discussed the problem of the expansion of the general bounded or almost bounded function in a trigonometrical form. It is manifest that expansions of this type throw much light on the properties of the difference equation

$$f(x+1)-f(x)=\phi(x). \tag{1}$$

It is the purpose of this paper to develop this method of treatment to the extent at least of duplicating certain results which Nörlund[2] has reached by other methods. In the course of this investigation, methods will be developed which have possible applications in fields much wider than the calculus of finite differences, such as the calculus of operations and the methods of Heaviside for the solution of the partial differential equations of the electric circuit. I shall be content to indicate these here, leaving their elaboration to a subsequent paper.[3]

We shall put

$$\Theta(x)=0; \qquad [0\leq x\leq \tfrac{1}{2}]$$

$$\Theta(x)=\int_{\frac{1}{2}}^{x} e^{\frac{1}{6\xi^3-7\xi+2}}\,d\xi; \qquad [\tfrac{1}{2}\leq x\leq \tfrac{2}{3}]$$

$$\Theta(x)=\int_{\frac{1}{2}}^{\frac{2}{3}} e^{\frac{1}{6\xi^3-7\xi+2}}\,d\xi. \qquad [\tfrac{2}{3}\leq x]$$

We shall write

$$\Phi(x)=\frac{1}{\pi}\int_0^\infty \left[1-\frac{\Theta(a)}{\Theta(1)}\right]\cos ax\,da$$

and

$$\Psi(x)=\frac{1}{\pi}\int_0^\infty \frac{\Theta(a)}{a\Theta(1)}\sin ax\,da.$$

[1] On the Representation of Functions by Trigonometrical Integrals, Mathematische Zeitschrift.

[2] N. E. Nörlund, Mémoire sur le calcul aux différences finies. Acta Mathematica, V. 44 (1923), pp. 71–212.

[3] The Operational Calculus, forthcoming in the Mathematische Annalen.

We shall then put

$$\phi_1(x) = \int_{-\infty}^{\infty} \phi(\xi)\Phi(\xi - x)d\xi$$

and

$$\phi_2(x) = \phi(x) - \phi_1(x).$$

If

$$f_1(x+1) - f_1(x) = \phi_1(x) \tag{2}$$

and

$$f_2(x+1) - f_2(x) = \phi_2(x), \tag{3}$$

we shall manifestly have

$$f_1(x+1) + f_2(x+1) - [f_1(x) + f_2(x)] = \phi_1(x) + \phi_2(x).$$

Nörlund points out that the difference equation (1) has the formal solutions

$$-\sum_{s=0}^{\infty} \phi(x+s)$$

and

$$\int_a^x \phi(z)dz + \sum_{\nu=1}^{\infty} \frac{B_\nu}{\nu!} \phi^{(\nu-1)}(x),$$

which, however, are generally divergent. In place of these, he develops the solution

$$\lim_{\eta \to 0} \left\{ \int_a^{\infty} \phi(x) e^{-\eta x} dx - \sum_{s=0}^{\infty} \phi(x+s) e^{-\eta(x+s)} \right\},$$

which converges whenever:

(1) The function $\phi(x)$ possesses for all positive values of x a continuous derivative of order m which tends to 0 as x increases indefinitely; and

(2) The integral

$$\int_0^{\infty} B_m(-z)\phi^{(m)}(x+\omega z)dz$$

converges uniformly over the interval (0, 1), where

$$B_{2m}(x) = (-1)^{m+1} 2(2m)! \sum_{k=1}^{\infty} \frac{\cos 2\pi kx}{(2\pi k)^{2m}};$$

$$B_{2m+1}(x) = (-1)^{m+1} 2(2m+1)! \sum_{k=1}^{\infty} \frac{\sin 2\pi kx}{(2\pi k)^{2m+1}}.$$

We shall show that under Nörlund's conditions on $\phi(x)$,

$$f_1(x) = \int_a^x \phi_1(z)dz + \sum_{\nu=1}^{\infty} \frac{B_\nu}{\nu!} \phi_1^{(\nu-1)}(x) \tag{4}$$

and

$$f_2(x) = -\sum_{s=0}^{\infty} \phi_2(x+s) \tag{5}$$

converge to functions satisfying respectively (2) and (3). We shall show, moreover, that the solution which we thus obtain is identically that of Nörlund.

The Properties of $\Phi(x)$ and $\Psi(x)$

2. $\Theta(\alpha)$ clearly possesses at every point derivatives of all orders. Thus integrating by parts,

$$\Phi(x) = \frac{\pm 1}{\pi \Theta(1) x^n} \int_0^1 \begin{matrix}\sin \\ \cos\end{matrix} \alpha x \, \frac{d^n \Theta(\alpha)}{d\alpha^n} d\alpha$$

and

$$\Psi(x) = \pm \frac{1}{\pi x^n} \int_1^{\infty} \frac{d^n}{d\alpha^n} \left[\frac{\Theta(\alpha)}{\alpha \Theta(1)}\right] \begin{matrix}\sin \\ \cos\end{matrix} \alpha x \, d\alpha.$$

We hence see that whatever k may be, we have at $\pm\infty$

$$\Phi(x) = o(x^{-k}); \quad \Psi(x) = o(x^{-k}).$$

Similarly, $\Phi^{(n)}(x)$ and

$$\Psi^{(-n)}(x) = (-1)^n \int_x^{\infty} d\xi_1 \int_{\xi_1}^{\infty} d\xi_2 \ldots \int_{\xi_{n-1}}^{\infty} \Psi(\xi_n) d\xi$$

exist, and are $o(x^{-k})$ at infinity for all finite k.

Now let $\phi(x)$ be any function of finite order at $\pm\infty$. We have

$$\begin{aligned}
\int_A^x \phi_1(x)dx &= \int_A^x du \int_{-\infty}^{\infty} \phi(\xi)\Phi(\xi-u)d\xi \\
&= \int_{-\infty}^{\infty} \phi(\xi)d\xi \int_A^x \Phi(\xi-u)du \\
&= \int_{-\infty}^{\infty} \phi(\xi)d\xi \int_{\xi-x}^{\xi-A} \Phi(u)du \\
&= \frac{1}{\pi}\int_{\infty}^{\infty} \phi(\xi)d\xi \int_{\xi-x}^{\xi-A} du \int_0^1 \left[1-\frac{\Theta(\alpha)}{\Theta(1)}\right]\cos \alpha u\, d\alpha \\
&= \frac{1}{\pi}\int_{-\infty}^{\infty} \phi(\xi)d\xi \int_0^1 \left[1-\frac{\Theta(\alpha)}{\Theta(1)}\right] d\alpha \int_{\xi-x}^{\xi-A} \cos \alpha u\, du \\
&= \frac{1}{\pi}\int_{-\infty}^{\infty} \phi(\xi)d\xi \int_0^{\infty} \left[1-\frac{\Theta(\alpha)}{\Theta(1)}\right] \\
&\qquad\qquad \frac{\sin \alpha(\xi-A)-\sin \alpha(\xi-x)}{\alpha}\, dx \\
&= \int_{-\infty}^{\infty} \phi(\xi)d\xi[\Psi(\xi-x)-\Psi(\xi-A)]+\int_A^x \phi(\xi)d\xi.
\end{aligned}$$

Thus by differentiation,

$$\phi_1(x)-\frac{d}{dx}\int_{\infty}^{\infty} \phi(\xi)\Psi(\xi-x)d\xi=\phi(x).$$

Hence if $\phi(x)$ admits a derivative which is of finite order at $\pm\infty$

$$\begin{aligned}
\phi_2(x) &= -\frac{d}{dx}\int_{\infty}^{\infty} \phi(\xi)\Psi(\xi-x)d\xi \\
&= -\frac{d}{dx}\int_{-\infty}^{\infty} \phi(\xi)d\Psi^{(-1)}(\xi-x) \\
&= \frac{d}{dx}\int_{-\infty}^{\infty} \phi'(\xi)\Psi^{(-1)}(\xi-x)d\xi \\
&= \lim_{h\to 0}\int_{-\infty}^{\infty} \phi'(\xi)\frac{\Psi^{(-1)}(\xi-x-h)-\Psi^{(-1)}(\xi-x)}{h}d\xi
\end{aligned}$$

We have already seen that $u^n\Psi(u)$ has the Fourier integral expansion

$$u^n\Psi(u)=\frac{1}{\pi}\int_0^\infty \frac{\Theta(\alpha)}{\alpha\Theta(1)}u^n\sin\alpha u\,d\alpha$$

$$=\pm\frac{1}{\pi}\int_0^\infty \frac{d^n}{d\alpha^n}\left[\frac{\Theta(\alpha)}{\alpha\Theta(1)}\right]\begin{matrix}\sin\\ \cos\end{matrix}\alpha u\,d\alpha.$$

In a similar manner, we may prove that

$$u^n\left[\frac{\Psi^{(-1)}(u+h)-\Psi^{(-1)}(u)}{h}\right]=\pm\frac{1}{\pi}\int_0^\infty \frac{d^n}{d\alpha^n}\left[\frac{\Theta(\alpha)\sin\frac{\alpha h}{2}}{\frac{\alpha^2 h}{2}\Theta(1)}\right]\begin{matrix}\sin\\ \cos\end{matrix}\left(u+\frac{h}{2}\right)d\alpha.$$

There is no difficulty in establishing that

$$\frac{d^n}{d\alpha^n}\left[\frac{\Theta(\alpha)}{\alpha\Theta(1)}\right]=\underset{h\to 0}{l.m.}\left[\frac{\Theta(\alpha)\sin\frac{\alpha h}{2}}{\frac{\alpha^2 h}{2}\Theta(1)}\right]$$

where the *l.m.* is understood to extend to the entire postive half-line. Hence

$$u^n\Psi(u)=\underset{h\to 0}{l.m.}\ u^n\left[\frac{\Psi^{(-1)}(u+h)-\Psi^{(-1)}(u)}{h}\right],$$

where the *l.m.* refers to the limit in the mean from $-\infty$ to ∞. Thus if $\phi'(x)$ is of finite order at $\pm\infty$, we have

$$\phi_2(x)=\lim_{h\to 0}\int_{-\infty}^{\infty}\phi'(\xi)\frac{\Psi^{(-1)}(\xi-x-h)-\Psi^{(-1)}(\xi-x)}{h}d\xi$$

$$=-\int_{-\infty}^{\infty}\phi'(\xi)\Psi(\xi-x)d\xi.$$

Integrating by parts, we see that if $\phi^{(m)}(x)$ is of finite order at $\pm\infty$,

$$\phi_2(x)=-\int_{-\infty}^{\infty}\phi^{(m)}(\xi)\,\Psi^{(-m+1)}(\xi-x)d\xi.$$

From the standpoint of trigonometric representation, $\phi_1(x)$ represents those components of $\phi(x)$ with a frequency less than π, together with a portion of those components with a frequency

between π and $4\pi/3$. On the other hand, $\phi_2(x)$ represents the remainder of those components with a frequency between π and $4\pi/3$, together with all those having a higher frequency.

The Low Frequency Components of $\phi(x)$.

3. Up to this point we have made no direct reference to the difference equation, and indeed our methods are useful in the treatment of a much wider class of problems concerning the operational method. We now turn to the difference equation (2) which we wish to prove to be satisfied by series (4), under the hypothesis that $\phi^{(m)}(x)$ is bounded.

We have

$$\phi_1^{(m+p)}(x) = (-1)^{(m+p)} \int_{-\infty}^{\infty} \phi(\xi) \Phi^{(m+p)}(\xi - x) d\xi$$
$$= (-1)^p \int_{-\infty}^{\infty} \phi^{(m)}(\xi) \Phi^{(p)}(\xi - x) d\xi.$$

Now,

$$u^2 \Phi^{(p)}(u) = \pm \frac{u^2}{\pi} \int_0^{\infty} \left[1 - \frac{\Theta(a)}{\Theta(1)}\right] a^p \begin{matrix} \sin \\ \cos \end{matrix} au \, da$$

$$= \pm \frac{1}{\pi} \int_0^{\infty} \frac{d^2}{da^2} \left\{ a^p \left[1 - \frac{\Theta(a)}{\Theta(1)}\right] \right\} \begin{matrix} \sin \\ \cos \end{matrix} au \, du$$

$$= \pm \frac{1}{\pi} \int_0^{\infty} \left\{ p(p-1) a^{p-2} \left[1 - \frac{\Theta(a)}{\Theta(1)}\right] - p a^{p-1} \frac{\Theta'(a)}{\Theta(1)} \right.$$
$$\left. - a^p \frac{\Theta''(a)}{\Theta(1)} \right\} \begin{matrix} \sin \\ \cos \end{matrix} au \, du$$

Hence there is a constant K independent of p such that

$$|\Phi^{(p)}(u)| \leq \frac{K(\frac{2}{3})^p}{u^2}.$$

Similarly, there is an L independent of p such that

$$|\Phi^{(p)}(u)| \leq L(\tfrac{2}{3})^p$$

and an M depending only on ϕ such that for all positive p,

$$\phi_1^{(m+p)}(x) \leq M(\tfrac{2}{3})^p.$$

The series

$$\sum_{\nu=1}^{\infty} \frac{B_\nu}{\nu!}\, \xi^\nu$$

has the radius of convergence 1. Hence

$$f_1(x) = \int_a^x \phi_1(z)dz + \sum_{\nu=1}^{\infty} \frac{B_\nu}{\nu!}\, \phi_1{}^{(\nu-1)}(x)$$

converges absolutely and uniformly for any finite range of x. Thus

$$f_1(x+1) - f_1(x) = \int_x^{x+1} \phi_1(z)dz + \sum_{\nu=1}^{\infty} \frac{B_\nu}{\nu!}\left[\phi_1{}^{(\nu-1)}(x+1) - \phi_1{}^{(\nu-1)}(x)\right]$$

$$= \int_x^{x+1} \phi_1(z)dz - \sum_{\nu=1}^{\infty} \frac{B_\nu}{\nu!}\left[\phi_1{}^{(\nu)}(x) + \frac{\phi_1{}^{(\nu+1)}(x)}{2!} + \ldots + \frac{\phi_1{}^{(\nu+k)}(x)}{k!} + \ldots\right]$$

$$= \int_x^{x+1} \phi_1(z)dz - \sum_{k=1}^{\infty} \sum_{\nu=1}^{k} \frac{B_\nu \phi_1{}^{(k)}(x)}{\nu!\, k!}$$

$$= \int_x^{x+1} \phi_1(z) - \sum_{k=1}^{\infty} \frac{\phi_1{}^{(k)}(x)}{(k+1)!} = \phi_1(x).$$

The inversion of limits in this reasoning is justified by the fact that the double series which is summed is absolutely convergent.

In a similar manner, we may show that the series

$$F(x) = \int_x^{-a} \Phi(z)dz + \sum_{\nu=1}^{\infty} \frac{B_\nu}{\nu!}(-1)^{\nu-1}\Phi^{(\nu-1)}(x)$$

is convergent, and solves the difference equation

$$F(x-1) - F(x) = \Phi(x).$$

It is easy to verify that

$$f_1(x) = \int_{-\infty}^{\infty} \phi(\xi)\left[F(\xi - x) - \int_{\xi-x}^{-a} \Phi(z)dz\right]d\xi + \int_a^x \phi_1(z)dz. \qquad (6)$$

The Fourier integral expression for

$$F(x)-\int_x^{-\infty}\Phi(z)dz$$

is

$$\sum_{\nu=1}^{\infty}\frac{B^\nu}{2\pi(\nu!)}\int_0^{\infty}\left[1-\frac{\Theta(\alpha)}{\Theta(1)}\right]\left[(-i\alpha)^{\nu-1}e^{i\alpha x}+(i\alpha\nu)^{\nu-1}e^{-i\alpha x}\right]d\alpha$$

$$=\frac{1}{2\pi}\int_0^{\infty}\left[1-\frac{\Theta(\alpha)}{\Theta(1)}\right]\left\{\frac{e^{i\alpha x}}{e^{-i\alpha}-1}+\frac{e^{i\alpha x}}{i\alpha}+\frac{e^{-i\alpha x}}{e^{i\alpha}-1}-\frac{e^{-i\alpha x}}{i\alpha}\right\}d\alpha$$

$$=\frac{1}{\pi}\int_0^{\infty}\left[1-\frac{\Theta(\alpha)}{\Theta(1)}\right]\left\{\frac{\sin\alpha x}{\alpha}+\frac{\cos(x+1)\alpha-\cos\alpha x}{2-2\cos\alpha}\right\}d\alpha.$$

Nörlund's principal solution of (1) is that for which there exists

$$\lim_{x\to\infty}f^{(m)}(x).$$

Let

$$\lim_{x\to\infty}\phi^{(m-1)}(x)=0.$$

Differentiating expression (6) m times, we get

$$f_1^{(m)}(x)=\int_{-\infty}^{\infty}\phi^{(m-1)}(\xi)\frac{d}{dx}\left[F(\xi-x)-\int_{-x}^{-\infty}\Phi(z)dz\right]d\xi+\phi_1^{(m-1)}(x).$$

It is then easy to see that the solution which we have obtained for (2) is the principal solution.

The High Frequency Components of $\phi(x)$

4. Let us now subject our function $\phi(x)$ to the entire force of Nörlund's hypothesis. We shall have formally

$$f_2(x)=\sum_{s=0}^{\infty}\int_{-\infty}^{\infty}\phi^{(m)}(\xi)\Psi^{(-m+1)}(\xi-x-s)d\xi.$$

We wish to prove the convergence of this series.

To begin with, let us consider the function

$$K(u)=\sum_{s=-\infty}^{\infty}\Psi^{(-m+1)}(u-s)$$

This function exists, and is periodic with period 1. It is hence appropriate to consider its Fourier series.

We have

$$\int_0^1 K(u)\sin 2\pi nu\,du = \sum_{s=-\infty}^{\infty}\int_0^1 \Psi^{(-m+1)}(u-s)\sin 2\pi nu\,du$$

$$= \sum_{s=-\infty}^{\infty}\int_{-s}^{1-s} \Psi^{(-m+1)}(\nu)\sin 2\pi n\nu\,d\nu$$

$$= \int_{-\infty}^{\infty} \Psi^{(-m+1)}(\nu)\sin 2\pi n\nu\,d\nu$$

$$= \pm\frac{1}{\pi}\int_{-\infty}^{\infty}\sin 2\pi n\nu\,d\nu \int_0^{\infty}\frac{\Theta(a)}{a^m\Theta(1)}\begin{matrix}\sin\\ \cos\end{matrix} a\nu\,da$$

$$\begin{cases} = \pm\dfrac{\Theta(2\pi n)}{(2\pi n)^m\Theta(1)} = \dfrac{\pm 1}{(2\pi n)^m}; & [m \text{ odd}] \\ = 0. & [m \text{ even}] \end{cases}$$

Again,

$$\int_0^1 K(u)\cos 2\pi nu\,du \begin{cases} = \pm\dfrac{1}{(2\pi n)^m}; & [m \text{ even}] \\ = 0. & [m \text{ odd}] \end{cases}$$

This means, however, that

$$K(u) = \frac{1}{n!}\dot{B}_n(u)$$

That is, Nörlund's condition tells us that

$$\int_{-\infty}^{\infty}\phi^m(\xi)\left[\sum_{-\infty}^{\infty}\Psi^{(-m+1)}(\xi-x-s)\right]d\xi$$

exists.

Now let us compare

$$U = \int_{\mu}^{\nu}\phi^{(m)}(\xi)\left[\sum_{-\infty}^{\infty}\Psi^{(-m+1)}(\xi-x-s)\right]d\xi$$

and

$$V = \int_{-\infty}^{\infty}\phi^{(m)}(\xi)\left[\sum_{\mu}^{\nu}\Psi^{(-m+1)}(\xi-x-s)\right]d\xi$$

We have

$$U-V=\int_{-\infty}^{\infty}\phi^{(m)}(\xi)Q(\xi-x)d\xi,$$

where

$$Q(u)=\sum_{s=\mu}^{\nu}\Psi^{(-m+1)}(u-s);\quad [u<\mu \text{ or } u>\nu)$$

$$Q(u)=\sum_{-\infty}^{\mu-1}\Psi^{(-m+1)}(u-s)+\sum_{\nu+1}^{\infty}\Psi^{(-m+1)}(u-s).\quad [\mu<u<\nu].$$

We have seen that

$$\Psi^{(-m+1)}(u)\leq\frac{R}{S+u^4}.$$

Thus

$$|Q(u)|\leq\sum_{t=0}^{\infty}\left[\frac{R}{S+[|u-\mu|+t]^4}+\frac{R}{S+[|u-\nu|+t]^4}\right]$$

$$\leq\frac{R_1}{S_1+|u-\mu|^3}+\frac{R_2}{S_2+|u-\nu|^3},$$

where R_1, R_2, S_1, and S_2 are properly chosen constants. If now $\phi^{(m)}(x)$ is a bounded function vanishing at $+\infty$, it follows that for any finite range of x, $U-V$ tends to 0 as μ and ν tend to $+\infty$ independently. Now, we already know that U tends to 0 as μ and ν tend to $+\infty$ independently. Thus if $\nu>\mu$, we have

$$0=\lim_{\mu\to\infty}\int_{-\infty}^{\infty}\phi^{(m)}(\xi)\left[\sum_{\mu}^{\nu}\Psi^{(-m+1)}(\xi-x-s)\right]d\xi$$

$$=\lim_{\mu\to\infty}\sum_{\mu}^{\nu}\int_{-\infty}^{\infty}\phi^{(m)}(\xi)\Psi^{(-m+1)}(\xi-x-s)d\xi.$$

Therefore

$$\sum_{s=0}^{\infty}\int_{-\infty}^{\infty}\phi^{(m)}(\xi)\Psi^{(-m+1)}(\xi-x-s)d\xi$$

converges. It manifestly satisfies equation (3), and as manifestly

is the principal solution. Thus the principal solution of (1) is

$$f_1(x)+f_2(x)=\int_a^x \phi_1(z)dz+\sum_{\nu=1}^{\infty}\frac{B_\nu}{\nu!}\phi_1^{(\nu-1)}(x)-\sum_{s=0}^{\infty}\phi_2(x+s)$$

$$=\int_a^x dz\int_{-\infty}^{\infty}\phi(\xi)\Phi(\xi-z)d\xi$$

$$+\frac{1}{\pi}\int_{-\infty}^{\infty}\phi(\xi)d\xi\int_0^{\infty}\left[1-\frac{\Theta(\alpha)}{\Theta(1)}\right]\left\{\frac{\sin\alpha(\xi-x)}{\alpha}-\frac{\sin\alpha(\xi-x+\frac{1}{2})}{2\sin\frac{\alpha}{2}}\right\}d\alpha$$

$$+\sum_{s=0}^{\infty}\int_{-\infty}^{\infty}\phi^{(m)}(\xi)\Psi^{(-m+1)}(\xi-x-s)d\xi.$$

NOTE ON A METHOD OF EVALUATING THE COMPLEX ROOTS OF SIXTH- AND HIGHER- ORDER EQUATIONS

By L. F. Woodruff

The study of oscillatory transients in complicated electrical and mechanical circuits involves the solution of linear differential equations of high order, and this in turn involves the solution of algebraic equations of like order, and having complex roots. The purpose of this note is to present a practical method of obtaining numerical values of the roots of sixth- and higher-order equations.

Let the general sixth-order equation with complex roots be

$$\boldsymbol{x}^6+a\boldsymbol{x}^5+b\boldsymbol{x}^4+c\boldsymbol{x}^3+d\boldsymbol{x}^2+e\boldsymbol{x}+f=0, \tag{1}$$

where the heavy or Clarendon type denotes that $\boldsymbol{x}$ is a complex number. Let its magnitude be x, and let θ be defined from the relation that $\boldsymbol{x}=x(\cos\theta+j\sin\theta)$. Since complex roots occur in pairs, three values of x and three values of θ will define the six roots.

If the seven terms of (1) are plotted, they may be regarded as seven vectors having the angle θ between each adjacent pair. The sum of the seven vectors must be zero if the equation is to be satisfied. Consider the components respectively perpendicular to and parallel to $c\boldsymbol{x}^3$.

$$(x^6-f)\sin 3\theta+(ax^5-ex)\sin 2\theta+(bx^4-dx^2)\sin\theta=0 \tag{2}$$

$$(x^6+f)\cos 3\theta+(ax^5+ex)\cos 2\theta+(bx^4+dx^2)\cos\theta+cx^3=0 \tag{3}$$

From (2)

$$(x^6-f)(3\sin\theta-4\sin^3\theta)+(ax^5-ex)2\sin\theta\cos\theta+(bx^4-dx^2)\sin\theta=0.$$

Whence, dividing out $\sin\theta$,

$$4(x^6-f)\cos^2\theta+2(ax^5-ex)\cos\theta-(x^6-bx^4+dx^2-f)=0. \tag{A}$$

From (3),

$$(x^6+f)(4\cos^3\theta-3\cos\theta)+(ax^5+ex)(2\cos^2\theta-1)+(bx^4+dx^2)\cos\theta+cx^3=0. \tag{4}$$

If equation (A) is multiplied by $\cos\theta$, and the terms containing

$\cos^3\theta$ then eliminated between this and equation (4), there results:

$$4(ex^7-afx^5)\cos^2\theta-2(x^{12}-dx^9+bfx^4-f^3)\cos\theta$$
$$-(ax^{11}-cx^9+ex^7-afx^5+cfx^3-efx)=0. \quad (5)$$

If now the terms containing $\cos^2\theta$ are eliminated between equations (A) and (5), there results

$$\cos\theta=\begin{cases} -ax^{17}+cx^{15}+(af-be)x^{11}+(de+abf-2cf)x^9+(ef-adf)x^7 \\ \qquad +cf^2x^3-ef^3x \\ \hline 2[x^{18}-dx^{14}-(f-ae)x^{12}+(bf-a^2f)x^{10}+(df-e^2)x^8 \\ \qquad +(aef-f^2)x^6-bf^2x^4+f^3] \end{cases} \quad (B)$$

Values of x and θ which will simultaneously satisfy (A) and (B) will furnish roots of the original sixth-order equation. These values may best be found by plotting the two curves, assuming values for x and computing the corresponding values of $\cos\theta$. Furthermore, $\cos\theta$ must be limited to values less in absolute value than unity.

Example. Find the roots of the equations

$$x^6-22x^5+10{,}521\,x^4-140{,}628\,x^3+3{,}920{,}868\,x^2+27{,}391{,}840\,x$$
$$+65{,}104{,}000=0.$$

Commencing with zero and advancing by units for values of x in (A) and (B), there results the following table

Assumed Value of x	*Cos θ from Eq. (A)*		*Cos θ from Eq. (B)*
0	+0.5	−0.5	0
1	+0.390857	−0.601207	−0.211462
2	+0.274078	−0.694782	−0.430451
3	+0.150660	−0.781756	−0.669013
4	+0.022510	−0.864132	−0.950270
5	−0.111644	−0.940876	
6	−0.232587	−1.031723	

A six-place calculating machine was used in making the calculations. There is evidently a crossing of the two curves between $x=3$ and $x=4$, and a straight-line interpolation or a plot shows that it is approximately at $x=3.6$. If now Newton's interpolation

formula is used to calculate the values of cos θ for this value of x, from (A) there results cos $\theta = 0.8316$, and from B, cos $\theta = 0.8307$. A straight line interpolation between the values of cos θ for $x=3.6$ and $x=4$ give for the point of crossing $x=3.604$ and cos $\theta = -0.8319$. (The true values are $x=3.60555$ and cos $\theta = -0.832051$. Approximately four-figure accuracy is thus obtained.) Two roots of equation (1) are hence $3.604(-0.8319 \pm j\sqrt{1-0.8319^2}) = -3.000 \pm j\ 2.000$.

The remaining four roots may be found in the same way, but it is shorter, after obtaining the first two roots, to reduce the order of the equation by a division, in this case by $x^2+6x+13$, which gives for a quotient

$$x^4-28x^3+10{,}676\ x^2-204{,}320\ x+5{,}008{,}000=0.$$

The four remaining roots may now be found from this equation by any method desired, *e.g.*, by Lyon's[1] method, of which the solution given here is an extension. The remaining roots are $x=10 \pm j\ 20$ and $x=4 \pm j\ 100$.

The same general method of attack may be used for the solution of eighth- and higher-order equations with complex roots. In the eighth-order equation

$$x^8+ax^7+bx^6+cx^5+dx^4+ex^3+fx^2+gx+h=0$$

for instance, the two conditions to be satisfied are that

$$dx^4+(cx^5+ex^3)\cos\theta+(bx^6+fx^2)\cos 2\theta+(ax^7+gx)\cos 3\theta + (x^8+h)\cos 4\theta=0$$

$$(cx^5-ex^3)\sin\theta+(bx^6-fx^2)\sin 2\theta+(ax^7-gx)\sin 3\theta + (x^8-h\sin)\,4\theta=0.$$

By substitutions and interpolation similar to those made in the solution of the sextic, values of θ and x which will simultaneously satisfy these two equations may be found.

[1]Note on a Method of Evaluating the Complex Roots of a Quartic Equation, by W. V. Lyon, Journal Mathematics and Physics, V. 3, No. 3, April 1924.

MINIMAL VARIETIES OF TWO AND THREE DIMENSIONS WHOSE ELEMENT OF ARC IS A PERFECT SQUARE

By C. L. E. Moore

Some years ago Phillips and Moore[1] discussed geometries in which the locus of a point equidistant from a given point is a linear space. Later Moore[2] discussed the application of this geometry to minimal developables. The method used in these papers is algebraic, and only in the last paper is there any reference to arc length. In the present paper we shall start with the element of arc and develop the properties of the geometry in any number of dimensions. Before discussing the geometry on the variety itself we shall try to find the real Euclidean space, in which such varieties can be located.

1. **Varieties of two dimensions.** There are only two kinds.

$$ds^2 = du^2 \tag{1}$$

$$ds^2 = v^2 du^2 \tag{2}$$

depending on whether the element of arc is the square of an exact differential or not.[3] The second is the element of arc of the minimal developable and in three dimensions the parametric equations of such a surface are

$$x^{(i)} = f^{(i)}(u) + vg^{(i)}(u) \tag{3}$$

where $\Sigma\,]g^{(i)}]^2 = 0$, $\Sigma f_u^{(i)} g^{(i)} = 0$ [f_u denoting the derivative with respect to u]. In space of higher dimensions the equations of the surface do not need to be of this simple type.

If ds is defined by (1) we have[4],

$$\Sigma[x_u^{(i)}]^2 = 1, \quad \Sigma x_u^{(i)} x_v^{(i)} = 0, \quad \Sigma[x_v^{(i)}]^2 = 0 \tag{4}$$

[1] An Algebra of Plane Projective Geometry, Proceedings of the American Academy of Arts and Sciences, Vol. 47, 1912. Theory of Linear Distance and Angle. Same journal, Vol. 48, 1912.

[2] Geometry Whose Element of Arc is a Linear Differential Form, etc. Same journal, Vol. 50, 1915.

[3] This excludes geometries with element of arc identically zero.

[4] The method used is mainly that employed by Darboux, Théorie génerale des surfaces, Vol. I, p. 149.

where the coördinates are $x^{(1)}$, $x^{(2)}$, $x^{(3)}$ and subscripts denote derivatives.

In the future we shall, however, omit the indices in cases where no ambiguity can arise. Differentiating (4) we have

$$\begin{aligned} \Sigma x_u x_{uu} = 0, \quad \Sigma x_u x_{uv} = 0, \quad \Sigma x_u x_{vv} = 0. \\ \Sigma x_v x_{uu} = 0, \quad \Sigma x_v x_{uv} = 0, \quad \Sigma x_v x_{vv} = 0. \end{aligned} \tag{5}$$

From (4) and (5) the set of equations

$$\Sigma \lambda x_v, \quad \Sigma \lambda x_{uv} = 0, \quad \Sigma \lambda x_{uv}, \quad \Sigma \lambda x_{vv} = 0 \tag{6}$$

are seen to be satisfied by x_u, x_v. Equations (6) are homogeneous, and since $x_u \neq K x_v$, the matrix

$$\| x_v x_{uu} x_{uv} x_{vv} \| \tag{7}$$

must be of rank one. This means that the coördinates of the surface must satisfy three linearly independent equations of the form

$$A_i x_{uu} + B_i x_{uv} + C_i x_{vv} + D_i x_v = 0. \tag{8}$$

It is evident that if any change of parameters is made, the coördinates must still satisfy three linear partial differential equations of the second order. Let us then take $x_1 = u$, $x_2 = v$, in which case (7) becomes

$$\begin{Vmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ x_u^{(3)} & x_1^{(3)} & x_{uu}^{(3)} & x_{uv}^{(3)} & x_{vv}^{(3)} \end{Vmatrix}$$

and since this matrix is of rank two we have

$$x_{uu}^{(3)} = x_{uv}^{(3)} = x_{vv}^{(3)} = 0$$

or

$$x^{(3)} = A x^{(1)} + B x^{(2)} + C$$

where A, B, C are constants.[5] The surface is therefore a plane and reference to (4) shows that it must be a minimal plane.

[5] It is easily seen that in any number of dimensions, if the coördinates satisfy three linear partial differential equations of the second order, the surface must necessarily be a plane. This is a slight extension of Segre's theorem. "If the coördinates of a surface satisfy two linear partial differential equations of the second order it must either lie in a 3-space or be the locus of tangent lines to a twisted curve." See Su una classa di superficie degl'iperspazi, Atti di Torino, 1907.

In 4-space equations (6) are still satisfied by x_u and x_v, but in this case (7) is of rank three and the coördinates satisfy only two equations of form (8) and by Segre's theorem[5] the surface must lie in a 3-space or be of the form

$$x = f(u) + vf_u(u)$$

and a simple calculation shows that the element of arc cannot be of the form desired. If the surface lies in a 3-space it must lie in a minimal 3-space, and it is not difficult to show that the only such surfaces are again minimal planes. Hence: *The only surfaces lying in a 3-space or a 4-space, whose element of arc is an exact differential, are minimal planes.*

In 5-space there are surfaces of the kind sought, among which are all the surfaces which lie in a minimal 3-space having a single line for absolute. Such a 3-space is

$$x^{(1)} + ix^{(2)} = 0, \qquad x^{(3)} + ix^{(4)} = 0.$$

and obviously any surface lying in it will have an exact differential for element of arc.

2. **Varieties of three dimensions.** A linear differential form in three variables, according to its class, can be put into one of the three forms,[6] corresponding to non-integrable, integrable and exact. We will then write the element of arc

$$ds^2 = (vdu + dt)^2 \tag{9}$$

$$ds^2 = (vdu)^2 \tag{10}$$

$$ds^2 = du^2 \tag{11}$$

If the element of arc is of form (9)

$$\begin{aligned} &\Sigma x_u^2 = v^2, \quad \Sigma x_u x_t = v, \quad \Sigma x_t^2 = 1 \\ &\Sigma x_u x_v = 0, \quad \Sigma x_v x_t = 0, \quad \Sigma x_v^2 = 0. \end{aligned} \tag{12}$$

Non-integrable case. By differentiation we have also

$$\Sigma x_u x_{vv} = 0, \quad \Sigma x_v x_{vv} = 0, \quad \Sigma x_t x_{vv} = 0. \tag{13}$$

The set of equations

$$\Sigma \lambda x_u = 0, \quad \Sigma \lambda x_v = 0, \quad \Sigma \lambda x_t = 0 \tag{14}$$

[5] E. von Weber, Vorlesungen über das Pfaff'sche Problem, page 149.

are satisfied by x_v, x_{vv}. In 4-space, then, x_v and x_{vv} must be proportional. Integrating these equations

$$x = \phi(u, t) + s\psi(u, t) \tag{15}$$

where s is a function of u, v, t. From (12), (13), ϕ and ψ must satisfy the following relations

$$\Sigma\phi_u\psi = 0, \quad \Sigma\phi_t\psi = 0, \quad \Sigma\psi^2 = 0, \quad \Sigma\psi_t^2 = 0$$
$$\Sigma\phi_t\psi_t = 0, \quad \Sigma\psi_{tt}\psi = 0, \quad \Sigma\psi_{tt}\psi_t = 0,$$
$$\Sigma\psi_u\psi_t = 0, \quad \Sigma\phi_{tt}\psi = 0.$$

Consequently equations

$$\Sigma\lambda\psi = 0, \quad \Sigma\lambda\psi_t = 0, \quad \Sigma\lambda\psi_{tt} = 0.$$

are satisfied by ψ, ψ_v, ψ_t.

Since ψ_u and ψ_t cannot be proportional there must be a linear relation connecting ψ, ψ_t, ψ_{tt}, and then there must be a second linear relation connecting these quantities or else a linear relation connecting ψ, ψ_u, ψ_t. In the latter case, by differentiating once with respect to t and once with respect to u, there will result three linear partial differential equations which ψ must satisfy and consequently ψ must be linear in u and t. Using the first of equations (16) we have

$$\psi = Au + Bt$$

where A, B are constants and

$$\Sigma A^2 = \Sigma B^2 = \Sigma AB = 0, \quad \Sigma\phi_t B = 0.$$

Making use of the fact that $\Sigma x_t^2 = 1$ and that the element of arc is a perfect square we obtain also

$$\Sigma\phi_u B = 0.$$

Integrating the last two equations we obtain

$$\Sigma B\phi = \text{const.}$$

Now multiplying equations (15) respectively by B_1, B_2, B_3, B_4 and summing we obtain

$$\Sigma Bx = K.$$

Hence the variety lies in a 3-space and must therefore be a 3-space. If now ψ, ψ_t, ψ_{tt} satisfy two linear relations we see that $\psi = k\psi_t$.

Integrating this we get

$$\psi^{(2)}=f_2(u)\psi_1, \quad \psi^{(3)}=f_3(u)\psi_1, \quad \psi^{(4)}=f_4(u)\psi_1.$$

Hence (15) can be written

$$x=\phi(u, t)+S\psi^{(1)}f(u)=\phi(u, t)+wf(u)$$

where $w=S\psi^{(1)}$.

We now have the following relations:

$$\Sigma\phi_u f=0, \quad \Sigma\phi_v f=0, \quad \Sigma\phi_u f_u=0, \quad \Sigma f^2=0, \quad \Sigma f_u{}^2=0,$$

and by the same argument as before we arrive at the conclusion

that $$f_u^{(1)}/f^{(1)}=f_u^{(2)}/f^{(2)}= \ \ldots$$

and consequently $$f_2=A^{(1)}f_1, \quad f_3=A^{(2)}f_1, \ \ldots$$

or $$x=\phi+wf_1A,$$

where $\Sigma A^2=0$ and again the variety must lie in 3-space. It is obvious that none of the other forms can exist in 4-space. Hence, *there is no variety of three dimensions having the element of arc equal to the square of a linear differential.*

In 5-space there are varieties having $ds^2=(vdu+dt)^2$ as conditions (12) can be satisfied.

Integrable case. For the integrable case the arc length has the form (10), and from the argument given above there can be no such variety in 4-space. In 5-space there are simple examples of such varieties.

$$x^{(1)}=\phi^{(1)}(u, v), \quad x^{(2)}=\phi^{(2)}(u, v), \quad x^{(3)}=\phi_3(u, v), \quad x^{(4)}=t, \quad x^{(5)}=it.$$

where $\Sigma\phi_u{}^2\Sigma\phi_v{}^2=(\Sigma\phi_u\phi_v)^2$. This variety lies in a minimal 4-space.

$$x=f(u)+vf_u\ (u)+tf_{uu}(u),$$

where $\Sigma f_u{}^2=0$, $\Sigma f_{uu}{}^2=0$. This is the locus of the osculating planes to a minimal curve satisfying the conditions given.

In this case we see that (14) are satisfied by x_v, x_t, x_{ut}, x_{vv}, x_{tt}, and hence the coördinates must satisfy four linear partial differential equations of the second order, and the property of these varieties has been given by Sisam.[7]

[7] Spreads which satisfy four or more homogeneous partial differential equations of the second order, American Journal, Vol. XXXIII.

Exact differential. If the element of arc has the form $ds^2=du^2$ it is easy to show that equations (14) have the solutions x_v, x_t, x_{uu}, x_{uv}, x_{vv}, x_{ut}, x_{vt}, x_{tt} and hence in a 5-space, since there cannot be more than two independent solutions, the coördinates must satisfy six linearly independent partial differential equations of second order. In 6-space (14) can have three independent solutions, and consequently the coördinates will satisfy five linear partial differential equations of second order. In Sisam's article it is shown that "The necessary and sufficient condition that a variety satisfy five linear partial differential equations of second order is that it be situated in a 4-space, or else be the locus of a singly infinite number of planes in which near-by planes intersect in lines." It is easily shown that in case the coördinates satisfy six such equations the variety must be a 3-space. In 6-space varieties of the kind mentioned must have one of the three forms

$$(a)\quad x=f(u)+vf_u(u)+tf_{uu}(u),$$
$$(b)\quad x=f(u)+vf_u(u)+kt,$$
$$(c)\quad x=f(u)+k_1v+k_2t,$$

where k_1, k_2, k_3 are constants. (a) will have the square of a perfect differential for arc length if $\Sigma f_u^2=0$, $\Sigma f_{uu}^2=0$; (b) will have if $\Sigma f_u^2=0$, $\Sigma kf_u=0$, $\Sigma k^2=0$; (c) will have if $\Sigma k_1^2=0$, $\Sigma k_2^2=0$, $\Sigma k_1k_2=0$, $\Sigma k_1f_u=\Sigma k_2f_u=0$. Varieties of this kind cannot lie in an ordinary 5-space, so if one does lie in a 5-space it must be minimal. An example is $x^{(1)}=u$, $x^{(2)}=iu$, $x^{(3)}=v$, $x^{(4)}=iv$, $x^{(5)}=f^{(5)}(t)$, $x^{(6)}=f^{(6)}(t)$.

3. **Geometry on Minimal Developables.** For convenience in discussing this geometry we will take the element of arc in the non-exact case as

$$ds=udv-vdu \tag{17}$$

With this choice of parameters we can easily demonstrate the following theorem: *The distance from* (u_1, v_1) to (u, v) *measured along the curve linear in* u, v *which passes through the two given points is given by the formula*

$$d=u_1v-uv_1. \tag{18}$$

We shall call this the *distance* between the two points and note

that it is a directed quantity. From (18) we see that the locus of points equidistant from (u_1, v_1) is linear in (u, v). From (17) it is evident that all the minimal lines on such a surface are linear in (u, v); *viz.*, the curves of the pencil $u = cv$, where c is a constant. The distance from all finite points to the point $u=0$, $v=0$ is zero. From (18), the distance between any two points on the infinite part of the surface is indeterminate.

Now let u, v be taken as the x, y-coördinates of a point in a euclidean plane. (Euclidean distance and angle serve only as a means of ordering the elements. We could have used the projective plane, but it seemed a little more convenient for the problem in hand to use the euclidean plane.) Then

$$ds = xdy - ydx,$$
$$d = x_1y_2 - x_2y_1.$$

The two exceptional elements then become the origin O, and the infinite line of the plane. The lines along which measurement is made are straight lines. The locus of points equidistant from $(x_1, y_1) = P$ lie on a line passing through the intersection of OP with the infinite line which gives a graphical construction for the locus. Then given O, the line at infinity and a *unit distance*, the measurement is all determined. Any linear transformation which leaves O and the line at infinity fixed will leave distance invariant.

Angle. The definition of distance which we have used is a specialization of ordinary distance, but such cannot be the case for angle since, for the surfaces here considered, euclidean angle becomes useless. We will then start with a new definition of angle. We first seek for an invariant of two lines which will have the principal properties of angle. From what has been said such an invariant is obtained by a correlation which preserves O and the line at infinity. Take for simplicity the correlation with respect to the unit circle having O for center. Then to the line $ux + vy = 1$. will correspond the point (u, v) and we will define the angle between two lines as the distance between their corresponding points through the correlation. Thus

$$\text{Angle} = u_1v_2 - u_2v_1.$$

The angle between two lines passing through O is infinite and the

angle between two lines intersecting on the infinite line is zero. From the definition angle has the following properties, (*a*) If the two lines coincide the angle between them is zero. (*b*) If the lines l, m, n meet in a point

$$\text{angle } (l, m)+\text{angle } (m, n)=\text{angle } (l, n).$$

(*c*) Angle is a directed quantity.

The triangle. Let the angles of the triangle be denoted by A, B, C, and the opposite sides by a, b, c. Then an easy calculation will give

$$\frac{A}{a}=\frac{B}{b}=\frac{C}{c}=\frac{a+b+c}{abc}\,\frac{ABc}{A+B+C}$$

or

$$A=\frac{a+b+c}{bc} \qquad a=\frac{A+B+C}{BC}, \quad \text{etc.}$$

Exact case. If the element of arc is a perfect differential the geometry is essentially that of a minimal plane. We will choose the parameters so that

$$ds=d\left(\frac{v}{u}\right),$$

and as above depict it on a reale uclidean plane where $x=u$, $y=v$. The distance between points, as before, will be measured along a curve linear in u, v. The locus of points equidistant from a given point will be a line passing through O. For finite points.

$$d=\frac{x_1y_2-x_2y_1}{x_1x_2}.$$

Then the lines through O are the minimal lines and the distance between two points on the y-axis is indeterminate. This then is a specialization of the preceding case in which the exceptional point lies on the exceptional line.

To define angle we will take a polar reciprocation with respect to a unit circle touching the y-axis at O, which preserves the origin and y-axis. This gives

$$\text{Angle}=y_2-y_1.$$

The triangle relations in this case become

$$\frac{A}{a}=\frac{B}{b}=\frac{C}{c}, \quad a+b+c=A+B+C=0.$$

4. **The geometry on the varieties of three dimensions.** *Non-Integrable case.* By a convenient choice of parameters we can write

$$ds = ydx - xdy + dz, \tag{20}$$

and again measuring distance along curves which are linear in these parameters

$$d = x_1y_2 - x_2y_1 + z_2 - z_1. \tag{21}$$

Hence the locus of points equally distant from a fixed point is represented by a linear equation in the parameters. Considering x_1, y_1, z_1, as coördinates in euclidean space and letting (20) define distance we see that the minimal lines are lines of a non-special linear complex. In this complex the polar plane of a point (x_1, y_1, z_1) will be the locus of points at zero distance from (x_1, y_1, z_1) or

$$xy_1 - x_1y + z - z_1 = 0.$$

The planes which represent the locus of points equidistant from (x_1, y_1, z_1) are therefore parallel to the plane of the complex passing through that point. The distance between points on the infinite plane is indeterminate. The absolute then consists of a linear line complex and the plane at infinity. The fact that the infinite plane is an exceptional element is due to the particular choice of parameters and was taken only for convenience.

Angle between two planes. Angle can again be defined as the dual of distance, but the duality must preserve the minimal lines and therefore will be the polar reciprocation in the complex itself. The angle between two planes will be defined as the distance between their poles with respect to the complex. The pole of the plane $ux + vy + wz = 1$ is $\left(\frac{v}{w}, \frac{-u}{w}, \frac{1}{w}\right)$, hence we have for the angle between two planes

$$A = \frac{u_2v_1 - u_1v_2 + w_1 - w_2}{w_1w_2}. \tag{22}$$

If either of the planes is parallel to OZ, the angle becomes infinite. The pole of the infinite plane is the intersection of OZ with this plane and this point has the same relation to the angle between two planes that the plane at infinity has to the distance between

points. If two planes intersect in a line of the complex the angle between them is zero.

Angle between two lines. The fundamental system now consists of a linear complex, the plane at infinity and its pole with respect to the complex. If we use the usual line coördinates in non-homogeneous form the complex is

$$p = p_{12} + p_{34} = 0$$

where $p_{12} = x_1y_2 - x_2y_1$, $p_{34} = z_1 - z_2$. The angle between two lines can then be expressed in terms of the fundamental point F, the fundamental plane V and the fundamental complex p. In terms of Grassmann products this is given by

$$\text{Angle } (rs) = \frac{(Fr.Vs)}{(pr)(ps)} . \tag{23}$$

The formula was derived by Phillips and Moore.[8]

In discussing the geometry on a variety the only important case of angle between lines is that between intersecting lines and in this case could have been defined by the distance between the corresponding points obtained by duality in their plane. Other invariants are given by Phillips and Moore.

Integrable case. The element of arc in this case can be written

$$ds = xdy - ydx$$

and the distance between two points is

$$d = x_1y_2 - x_2y_1$$

In this case the minimal lines become the special linear complex of lines which cuts the z-axis. The fundamental system in this case consists of a special complex, the plane at infinity and the point where the axis of the complex cuts the plane at infinity. The locus of points equidistant from a given point is a plane parallel to OZ, and cutting the infinite plane in the line in which the plane determined by the point and OZ cuts the infinite plane. This geometry then becomes the same as the case of two dimensions

[8] Linear distance and angle. Proceedings American Acad. Arts and Sciences. Vol. 48, 1912.

for the non-exact case, where we consider planes through OZ as the same as lines through O in the plane case.

Exact case. The element of arc is

$$ds = d\left(\frac{y}{x}\right).$$

The minimal lines are again the special linear complex as above. The distance between two points is

$$d = \frac{x_1y_2 - x_2y_1}{x_1x_2}.$$

The fundamental plane now becomes the yz-plane, and we again have essentially the same kind of geometry as for the plane exact case.

We can sum up by saying that we are led to three kinds of geometry.

(*a*) Non-Integrable. Fundamental system consists of a general linear complex, a plane and the pole of the plane with respect to the complex.

(*b*) Integrable. The fundamental system consists of a special linear complex, a plane and the pole of the plane, *i.e.*, the point in which the axis of the complex cuts the plane.

(*c*) Exact. The fundamental system same as (*b*) except that the axis of the fundamental complex lies in the fundamental plane.

In (*a*) there is no plane containing more than a single pencil of minimal lines while in (*b*) and (*c*) there is a pencil of planes passing through OZ in which every line is a minimal line.

We can now state the following theorem. If the element of arc of a variety of three dimensions is a perfect square, it contains a four parametered set of curves having the following property, — in general two points determine a curve, all the curves passing through any point generate a surface such that a curve of the system passing through two points lies wholly in the surface. The minimal curves belong to this system and in general a pencil of them lies in a surface of the system. Each such surface is the locus of points equidistant from a fixed point when distance is measured along lines of the system. In cases (*b*) and (*c*) there

are surfaces determined by two of the curves such that every curve of the system lying in them is a minimal curve.

In this geometry there is no such thing as minimal distance, but from the above discussion we see that the system of curves linear in the parameters have many of the properties of geodesic lines and the surfaces determined by two of them take the place of geodesic surfaces.

THE MULTIPLE COMPLEMENT OF ONE OR MORE POLYADICS

By Frank L. Hitchcock and Lepine Hall Rice

In this paper we shall define the multiple complement of a polyadic or a number of polyadics in any number of dimensions and shall express it in determinantal form.

The multiple complement is a generalization of Gibbs's double cross product in three dimensions, and also of the " space complement " of Naess.

The " space complement " of a vector or a number of vectors (taken in a definite order), say of k vectors, in n dimensions, was introduced by Naess,[1] who defined it by means of a determinant of order n having the n fundamental units for the elements of each of its first $n-k$ (identical) rows, k being not greater than n, and having the scalar components of the vectors for its last k rows. The fundamental units being $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n$, and the vectors in the order taken, being $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_k$, with

$$\mathbf{a}_j \equiv \mathbf{e}_1 a_{j1} + \mathbf{e}_2 a_{j2} + \ldots + \mathbf{e}_n a_{jn},$$

Naess's function is the $(n-k)$-adic

$$N \equiv \begin{vmatrix} \mathbf{e}_1 & \ldots & \mathbf{e}_n \\ \cdots & \cdots & \cdots \\ \mathbf{e}_1 & \ldots & \mathbf{e}_n \\ a_{11} & \ldots & a_{1n} \\ \cdots & \cdots & \cdots \\ a_{k1} & \ldots & a_{kn} \end{vmatrix} \equiv \sum_{i,j} (\pm_{i_1 i_2 \ldots i_{n-k} j_1 j_2 \ldots j_k}) \mathbf{e}_{i_1} \ldots \mathbf{e}_{i_{kn-}} a_{1j_1} \ldots a_{kj_k}, \tag{1}$$

where the sign-factor is Cayley's symbol meaning + or − according as the appended permutation is even or odd (and meaning zero if any two of the appended numbers are equal).

We prefer to express the sign factor in this way rather than to

[1] *On a special polyadic of order n−p which can be derived from any p independent vectors in an n-dimensional space and which can be regarded as a generalization of the vector product*, by Almar Naess, Videnskapsselskapets Skrifter. I. Mat.-naturv. Klasse. 1922. No. 13.

write it as $(-1)^\theta$ and attempt to evaluate Θ; the subsequent development of the subject is thereby greatly simplified.

In particular, the complement of k vectors each of one term

$$\mathbf{e}_{j_1} a_{1j_1},\ \mathbf{e}_{j_2} a_{2j_2},\ \ldots,\ \mathbf{e}_{j_k} a_{kj_k}$$

$$\text{is } \sum_i (\pm i_1 i_2 \ldots i_{n-k} j_1 j_2 \ldots j_k)\mathbf{e}_{i_1}\mathbf{e}_{i_2}\ldots\mathbf{e}_{i_{n-k}}\, a_{1j_1} a_{2j_2}\ldots a_{kj_k}, \tag{2}$$

and the complement of k fundamental unit vectors

$$\mathbf{e}_{j_1},\ \mathbf{e}_{j_2},\ \ldots,\ \mathbf{e}_{j_k}$$

$$\text{is } \sum_i (\pm i_1 i_2 \ldots i_{n-k} j_1 j_2 \ldots j_k)\mathbf{e}_{i_1}\mathbf{e}_{i_2}\ldots\mathbf{e}_{i_{n-k}}. \tag{3}$$

The latter result may be concisely expressed in words if we use the notation

$$\mathbf{E}_n \equiv \begin{vmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \ldots & \mathbf{e}_n \\ \mathbf{e}_1 & \mathbf{e}_2 & \ldots & \mathbf{e}_n \\ . & . & . & . \\ \mathbf{e}_1 & \mathbf{e}_2 & \ldots & \mathbf{e}_n \end{vmatrix}; \tag{4}$$

we may say that *the complement of k fundamental unit vectors is the cofactor in $\mathbf{E}_n$ of these vectors in the last k rows of $\mathbf{E}_n$.*

As already implied, the complement of k fundamental unit vectors vanishes if two of them are alike; that of k vectors each of one term vanishes if two of them contain the same $\mathbf{e}$; and that of any k vectors vanishes if no term can be formed in N.

We shall use the notation

$$N \equiv ((\mathbf{a}_1\mathbf{a}_2\ldots\mathbf{a}_k))_F^{(n)}, \tag{5}$$

and shall say, on occasion, that the k vectors are *folded* together in forming their complement.

As an example of (3),

$$((\mathbf{e}_4\mathbf{e}_1))_F^{(5)} \equiv \begin{vmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \mathbf{e}_3 & \mathbf{e}_4 & \mathbf{e}_5 \\ \mathbf{e}_1 & \mathbf{e}_2 & \mathbf{e}_3 & \mathbf{e}_4 & \mathbf{e}_5 \\ \mathbf{e}_1 & \mathbf{e}_2 & \mathbf{e}_3 & \mathbf{e}_4 & \mathbf{e}_5 \\ . & . & . & 1 & . \\ 1 & . & . & . & . \end{vmatrix} \equiv \sum_i (\pm i_1 i_2 i_3 41)\mathbf{e}_{i_1}\mathbf{e}_{i_2}\mathbf{e}_{i_3}. \tag{6}$$

The conception due to Gibbs of a double product of two dyadics extends to define the *multiple product* of a number of polyadics. In the formation of the p-tuple product of k p-ads

$$\mathbf{a}_1\mathbf{a}_2 \ldots \mathbf{a}_p, \quad \mathbf{b}_1\mathbf{b}_2 \ldots \mathbf{b}_p, \quad \ldots, \quad \mathbf{g}_1\mathbf{g}_2 \ldots \mathbf{g}_p \tag{7}$$

the k vectors $\mathbf{a}_1, \mathbf{b}_1, \ldots \mathbf{g}_1$ standing first in the p-ads go together and stand first in the product; the k vectors $\mathbf{a}_2, \mathbf{b}_2, \ldots, \mathbf{g}_2$ standing second in the p-ads go together and stand second in the product; and so on to the k vectors $\mathbf{a}_p, \mathbf{b}_p, \ldots, \mathbf{g}_p$ standing last in the p-ads. The manner of combining the vectors that go together depends on the specific character of the product that is to be made.[2] In a *multiple indeterminate* product these vectors form k-ads $\mathbf{a}_1\mathbf{b}_1 \ldots \mathbf{g}_1$, etc., and the result is a (kp)-ad. In what we shall call the *multiple complement* of the k p-ads the k vectors of each of the p sets are folded together into their complement so that the result is the indeterminate product of p successive $(n-k)$-adics

$$((\mathbf{a}_1\mathbf{b}_1 \ldots \mathbf{g}_1))_F^{(n)} \ ((\mathbf{a}_2\mathbf{b}_2 \ldots \mathbf{g}_2))_F^{(n)} \ \ldots ((\mathbf{a}_p\mathbf{b}_p \ldots \mathbf{g}_p))_F^{(n)}, \tag{8}$$

which is an $[(n-k)p]$-adic. In any case the multiple product of a number of polyadics will be defined as the sum of the multiple products of their distributed polyads.

Let us first find the multiple complement of k fundamental unit p-ads. Our notation here may seem less simple than the case admits, but is adopted in order to utilize the results without change in later steps of the work. We let

$$\mathbf{e}_{j_r(s)} \tag{9}$$

denote the fundamental unit occupying the rth place in the sth p-ad, whence the desired multiple complement, by a natural extension of the notation (5), may be denoted by

[2] Multiple dot products have been treated in a paper *On double polyadics, with application to the linear matrix equation*, by Frank L. Hitchcock, Proc. American Academy of Arts and Sciences, Vol. 58, No. 10, May, 1923.

$$\begin{aligned}&((\mathbf{e}_{j_1(1)}\mathbf{e}_{j_2(1)}\cdots\mathbf{e}_{j_p(1)},\\&\ \mathbf{e}_{j_1(2)}\mathbf{e}_{j_2(2)}\cdots\mathbf{e}_{j_p(2)},\\&\ \cdots\cdots\cdots\\&\ \mathbf{e}_{j_1(k)}\mathbf{e}_{j_2(k)}\cdots\mathbf{e}_{j_p(k)}))_F^{(n)}\end{aligned}\tag{10}$$

$$=\begin{cases}((\mathbf{e}_{j_1(1)}\mathbf{e}_{j_1(2)}\cdots\mathbf{e}_{j_1(k)}))_F^{(n)}\\((\mathbf{e}_{j_2(1)}\mathbf{e}_{j_2(2)}\cdots\mathbf{e}_{j_2(k)}))_F^{(n)}\\\cdots\cdots\cdots\\((\mathbf{e}_{j_p(1)}\mathbf{e}_{j_p(2)}\cdots\mathbf{e}_{j_p(k)}))_F^{(n)}\end{cases}\tag{11}$$

$$=\sum_i\begin{cases}(\pm i_1(1)i_1(2)\cdots i_1(n-k)j_1(1)j_1(2)\cdots j_1(k))\\(\pm i_2(1)i_2(2)\cdots i_2(n-k)j_2(1)j_2(2)\cdots j_2(k))\\\cdots\\(\pm i_p(1)i_p(2)\cdots i_p(n-k)j_p(1)j_p(2)\cdots j_p(k))\\\mathbf{e}_{i_1(1)}\mathbf{e}_{i_1(2)}\cdots\mathbf{e}_{i_1(n-k)}\\\mathbf{e}_{i_2(1)}\mathbf{e}_{i_2(2)}\cdots\mathbf{e}_{i_2(n-k)}\\\cdots\\\mathbf{e}_{i_p(1)}\mathbf{e}_{i_p(2)}\cdots\mathbf{e}_{i_p(n-k)}\end{cases}\tag{12}$$

This result may be concisely expressed in words if we introduce a determinantal notation analogous to that employed above for the complement of k fundamental unit vectors. We now set up a $(p+1)$-way matrix of order n whose $(p$-way) layers of the first direction are identical, each being composed of all the fundamental unit p-ads, the p-ad $\mathbf{e}_{\epsilon_1}\mathbf{e}_{\epsilon_2}\cdots\mathbf{e}_{\epsilon_p}$ occupying the place $\epsilon_1\epsilon_2\cdots\epsilon_p$ in the layer; this matrix may be written

$$E_{p+1}=\|\,\mathbf{e}_{\epsilon_1}\mathbf{e}_{\epsilon_2}\cdots\mathbf{e}_{\epsilon_p}\,|\,\mathbf{e}_{\epsilon_1}\mathbf{e}_{\epsilon_2}\cdots\mathbf{e}_{\epsilon_p}\,|\cdots|\,\mathbf{e}_{\epsilon_1}\mathbf{e}_{\epsilon_2}\cdots\mathbf{e}_{\epsilon_p}\,\|_{p+1}^{(n)}.\tag{13}$$

We now define the *leading primordinal multiple determinant of* E_{p+1}. By a *leading* determinant of a matrix of any number of directions we mean a determinant in which all directions are signant[3] if the matrix has an even number of directions, and in which all directions but the first are signant if the matrix has an odd number of directions. By a *primordinal* determinant of a

[3] Signancy is treated in a paper on *P-way determinants*, by Lepine Hall Rice, Amer. Journal of Math. 40, 1918, p. 242.

matrix we mean a determinant in each of whose terms an element from the first layer of the first direction is placed first, an element from the second layer is placed second, and so on. By a *multiple* determinant of a matrix we mean a determinant in each of whose terms the product of the elements is a multiple product, and, if the word multiple is not further qualified, a multiple indeterminate product. Now consider separately the first $n-k$ and the last k layers of E_{p+1}. Denote the general element in the place $i_1 i_2 \ldots i_p$ in the rth of the first $n-k$ layers by

$$\mathbf{e}_{i_1(r)}\mathbf{e}_{i_2(r)} \cdot \cdot \cdot \mathbf{e}_{i_p(r)} \tag{14}$$

and the general element in the place $j_1 j_2 \ldots j_p$ in the sth of the last k layers by

$$\mathbf{e}_{j_1(s)}\mathbf{e}_{j_2(s)} \cdot \cdot \cdot \mathbf{e}_{j_p(s)}. \tag{15}$$

Denote the leading primordinal multiple determinant of E_{p+1} by $\mathbf{E}_{p+1}$. Then in (10) we have a succession of k p-ads lying respectively in the last k layers of E_{p+1} and constituting the last k elements of a number of terms of $\mathbf{E}_{p+1}$, the first $n-k$ elements of each of these terms constituting a transversal of the $(p+1)$-way $(n-k)$-order minor of $\mathbf{E}_{p+1}$ complementary to these k elements (*i.e.*, lying in the $n-k$ layers, in each of the $p+1$ directions in E_{p+1}, in which these k elements do not lie), and the totality of these transversals being all of the transversals of this complementary minor. In (12) we have the expansion of this complementary minor considered as itself a leading primordinal multiple determinant and provided with the sign (contributed by the j's in the Cayley symbols) that gives us finally the signed minor which is the cofactor of the k p-ads of (10) in $\mathbf{E}_{p+1}$. Our verbal expression now is this:

The multiple complement of k fundamental p-ads is the cofactor in $\mathbf{E}_{p+1}$ of these p-ads in the last k layers of the first direction in $\mathbf{E}_{p+1}$.

As an example the multiple complement of the three fundamental unit hexads in five dimensions

$$\mathbf{e}_1\mathbf{e}_1\mathbf{e}_4\mathbf{e}_5\mathbf{e}_2\mathbf{e}_3, \quad \mathbf{e}_3\mathbf{e}_2\mathbf{e}_5\mathbf{e}_4\mathbf{e}_5\mathbf{e}_4, \quad \mathbf{e}_5\mathbf{e}_5\mathbf{e}_1\mathbf{e}_1\mathbf{e}_1\mathbf{e}_1 \tag{16}$$

$$\text{is} \sum_i (\pm i_1(1) i_1(2)\ 135) \cdot \cdot \cdot (\pm i_4(1) i_6(2)\ 341)\, \mathbf{e}_{i_1(1)}\mathbf{e}_{i_1(2)} \cdot \cdot \cdot \mathbf{e}_{i_6(1)}\mathbf{e}_{i_6(2)}, \tag{17}$$

a bihexadic of $2^6=64$ terms in which one term is

$$-\mathbf{e}_2\mathbf{e}_4\mathbf{e}_3\mathbf{e}_4\mathbf{e}_2\mathbf{e}_3\mathbf{e}_2\mathbf{e}_3\mathbf{e}_4\mathbf{e}_2\mathbf{e}_5,$$

the negative sign being obtained from

$$(\pm_{24135})(\pm_{24125})(\pm_{23431})(\pm_{23541})(\pm_{24251})(\pm_{25341})=-+--++=-;$$

and the remaining terms may all be obtained by permuting the two $\mathbf{e}$'s in one or more of the six pairs of $\mathbf{e}$'s in every possible way, the sign changing with each permutation.

It is now easy to form the multiple complement of k p-adics

$$\begin{aligned}
\mathbf{A}&\equiv\sum\mathbf{e}_{j_1(1)}\mathbf{e}_{j_2(1)}\cdot\cdot\cdot\mathbf{e}_{j_p(1)}a_{j_1(1)j_2(1)\cdot\cdot\cdot j_p(1)},\\
\mathbf{B}&\equiv\sum\mathbf{e}_{j_1(2)}\mathbf{e}_{j_2(2)}\cdot\cdot\cdot\mathbf{e}_{j_p(2)}b_{j_1(2)j_2(2)\cdot\cdot\cdot j_p(2)},\\
&\cdot\cdot\cdot\cdot\cdot\cdot\cdot\cdot\cdot\cdot\cdot\cdot\cdot\cdot\\
\mathbf{G}&\equiv\sum\mathbf{e}_{j_1(k)}\mathbf{e}_{j_2(k)}\cdot\cdot\cdot\mathbf{e}_{j_p(k)}g_{j_1(k)j_2(k)\cdot\cdot\cdot j_p(k)};
\end{aligned}\tag{18}$$

it will take the form of (8) with coefficients annexed and the summation made with respect to j as well as i, *viz.*,

$$((\mathbf{A},\mathbf{B},\ldots\mathbf{G}))_F^{(n)}=\sum_{i,j}\begin{cases}
(\pm i_1(1)i_1(2)\ldots i_1(n-k)j_1(1)j_1(2)\ldots j_1(k))\\
(\pm i_2(1)i_2(2)\ldots i_2(n-k)j_2(1)j_2(2)\ldots j_2(k))\\
\cdot\cdot\cdot\\
(\pm i_p(1)i_p(2)\ldots i_p(n-k)j_p(1)j_p(2)\ldots j_p(k))\\
\mathbf{e}_{i_1(1)}\mathbf{e}_{i_1(2)}\cdot\cdot\cdot\mathbf{e}_{i_1(n-k)}\\
\mathbf{e}_{i_2(1)}\mathbf{e}_{i_2(2)}\cdot\cdot\cdot\mathbf{e}_{i_2(n-k)}\\
\cdot\cdot\cdot\\
\mathbf{e}_{i_p(1)}\mathbf{e}_{i_p(2)}\cdot\cdot\cdot\mathbf{e}_{i_p(n-k)}\\
a_{j_1(1)j_2(1)\ldots j_p(1)}\\
b_{j_1(2)j_2(2)\ldots j_p(2)}\\
\cdot\cdot\cdot\\
g_{j_1(k)j_2(k)\ldots j_p(k)}
\end{cases}\tag{19}$$

This admits of very simple determinantal expression. We set up a $(p+1)$-way matrix M of order n by taking as each of the first $n-k$ p-way layers of the first direction the matrix of funda-

mental unit p-ads, and as the remaining k layers the k matrices of the a's, the b's, and so on up to the g's, so that

$$M=\left\|\mathbf{e}_{i_1(1)}\ldots\mathbf{e}_{i_p(1)}\right|\ldots\left|\mathbf{e}_{i_1(n-k)}\ldots\mathbf{e}_{i_p(n-k)}\left|a_{j_1(1)}\ldots j_p(1)\right|\right.\\ \left.\ldots\left|g_{j_1(k)}\ldots j_p(k)\right|\right\|_{p+1}^{(n)}. \tag{20}$$

We can now state the following theorem, whose truth is evident upon inspection of (19):

Theorem. The multiple complement of k p-adics in n dimensions, the p-adics being each reduced to a sum of fundamental p-ads, is equal to the leading primordinal multiple determinant of the $(p+1)$-way matrix of order n whose first $n-k$ p-way layers of the first direction are each the matrix of fundamental unit p-ads and whose last k p-way layers of the first direction are the matrices of coefficients of the k p-adics.

The *multiple complement of a single polyad* is the direct product of the complements of its vectors. This is in conformity with (8), which will then become

$$((\mathbf{a}_1))_F^{(n)}\quad(\mathbf{a}_2))_F^{(n)}\quad\ldots((\mathbf{a}_p))_F^{(n)}. \tag{21}$$

In (12), $n-k$ will be $n-1$, as also in (19), where all italic letters beyond a will disappear. If the rth vector in the p-ad be

$$\mathbf{e}_1a_{r1}+\mathbf{e}_2a_{r2}+\ldots+\mathbf{e}_na_{rn}, \tag{22}$$

then evidently

$$\mathbf{a}_1\mathbf{a}_2\ldots\mathbf{a}_p=\sum\mathbf{e}_{j_1}\mathbf{e}_{j_2}\ldots\mathbf{e}_{j_p}a_{1j_1}a_{2j_2}\ldots a_{pj_p}, \tag{23}$$

and (20) takes the form

$$\left\|\mathbf{e}_{i_1(1)}\ldots\mathbf{e}_{i_p(1)}\right|\ldots\left|\mathbf{e}_{i_1(n-1)}\ldots\mathbf{e}_{i_p(n-1)}\left|a_{1j_1}a_{2j_2}\ldots a_{pj_p}\right|\right\|_{p+1}^{(n)}; \tag{24}$$

that is, each of the elements of the last layer is the product of p coefficients one belonging to each vector of the p-ad, the p indices in the locant $j_1j_2\ldots j_p$ of the element in the layer determining the respective coefficients.

The *multiple complement of a polyadic* **A** will therefore have the matrix

$$\left\|\mathbf{e}_{i_1(1)}\ldots\mathbf{e}_{i_p(1)}\right|\ldots\left|\mathbf{e}_{i_1(n-1)}\ldots\mathbf{e}_{i_p(n-1)}\left|a_{j_1j_2}\ldots j_p\right|\right\|_{p+1}^{(n)}; \tag{25}$$

that is, the above theorem holds when $k=1$ as well as when $k>1$.

The multiple complement of n polyadics is evidently a scalar, the leading determinant of the matrix of their coefficients.

If the polyadics are vectors, the multiple complement has no multiple character but is simply Naess's complement.

If the polyadics are dyadics, their coefficients fill the last k strata of a cubic matrix of order n, the first $n-k$ strata being identical matrices of fundamental unit dyads. If $n=3$ and the dyadics are $\mathbf{A}_2^{(3)}$ and $\mathbf{B}_2^{(3)}$ we have, using $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$ for the fundamental unit vectors.

$$((\mathbf{A}_2^{(3)},\mathbf{B}_2^{(3)}))_F^{(3)} \equiv \left|\begin{array}{ccc|ccc|ccc} \mathbf{ii} & \mathbf{ij} & \mathbf{ik} & a_{11} & a_{12} & a_{13} & b_{11} & b_{12} & b_{13} \\ \mathbf{ji} & \mathbf{jj} & \mathbf{jk} & a_{21} & a_{22} & a_{23} & b_{21} & b_{22} & b_{23} \\ \mathbf{ki} & \mathbf{kj} & \mathbf{kk} & a_{31} & a_{32} & a_{33} & b_{31} & b_{32} & b_{33} \end{array}\right| . \quad (26)$$

This latter form is the same as that obtained for the double cross product of two dyadics in three dimensions by one of the present writers in a recent paper.[4] More generally, for any two p-adics in three dimensions, the determinant described in our theorem above is the same as the determinant given in the paper referred to, as an expression of the p-tuple cross product of the two p-adics.

In four dimensions the double complement of two dyadics $\mathbf{A}_2^{(4)}$ and $\mathbf{B}_2^{(4)}$ may be obtained from the 3-way matrix of the fourth order

$$\left\|\begin{array}{cccc|cccc} \mathbf{e}_1\mathbf{e}_1 & \mathbf{e}_1\mathbf{e}_2 & \mathbf{e}_1\mathbf{e}_3 & \mathbf{e}_1\mathbf{e}_4 & \mathbf{e}_1\mathbf{e}_1 & \mathbf{e}_1\mathbf{e}_2 & \mathbf{e}_1\mathbf{e}_3 & \mathbf{e}_1\mathbf{e}_4 \\ \mathbf{e}_2\mathbf{e}_1 & \mathbf{e}_2\mathbf{e}_2 & \mathbf{e}_2\mathbf{e}_3 & \mathbf{e}_2\mathbf{e}_4 & \mathbf{e}_2\mathbf{e}_1 & \mathbf{e}_2\mathbf{e}_2 & \mathbf{e}_2\mathbf{e}_3 & \mathbf{e}_2\mathbf{e}_4 \\ \mathbf{e}_3\mathbf{e}_1 & \mathbf{e}_3\mathbf{e}_2 & \mathbf{e}_3\mathbf{e}_3 & \mathbf{e}_3\mathbf{e}_4 & \mathbf{e}_3\mathbf{e}_1 & \mathbf{e}_3\mathbf{e}_2 & \mathbf{e}_3\mathbf{e}_3 & \mathbf{e}_3\mathbf{e}_4 \\ \mathbf{e}_4\mathbf{e}_1 & \mathbf{e}_4\mathbf{e}_2 & \mathbf{e}_4\mathbf{e}_3 & \mathbf{e}_4\mathbf{e}_4 & \mathbf{e}_4\mathbf{e}_1 & \mathbf{e}_4\mathbf{e}_2 & \mathbf{e}_4\mathbf{e}_3 & \mathbf{e}_4\mathbf{e}_4 \end{array}\right.$$

$$\left.\begin{array}{cccc|cccc} a_{11} & a_{12} & a_{13} & a_{14} & b_{11} & b_{12} & b_{13} & b_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} & b_{21} & b_{22} & b_{23} & b_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} & b_{31} & b_{32} & b_{33} & b_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} & b_{41} & b_{42} & b_{43} & b_{44} \end{array}\right\| \quad (27)$$

A Laplace expansion may be used if desired, giving

$$\sum (\pm_{\beta_1\beta_2\beta_3\beta_4})(\pm_{\gamma_1\gamma_2\gamma_3\gamma_4}) \left|\begin{array}{cc|cc} \mathbf{e}_{\beta_1}\mathbf{e}_{\gamma_1} & \mathbf{e}_{\beta_1}\mathbf{e}_{\gamma_2} & \mathbf{e}_{\beta_1}\mathbf{e}_{\gamma_1} & \mathbf{e}_{\beta_1}\mathbf{e}_{\gamma_2} \\ \mathbf{e}_{\beta_2}\mathbf{e}_{\gamma_1} & \mathbf{e}_{\beta_2}\mathbf{e}_{\gamma_2} & \mathbf{e}_{\beta_2}\mathbf{e}_{\gamma_1} & \mathbf{e}_{\beta_2}\mathbf{e}_{\gamma_2} \end{array}\right| \cdot \left|\begin{array}{cc|cc} a_{\beta_3\gamma_3} & a_{\beta_3\gamma_4} & b_{\beta_3\gamma_3} & b_{\beta_3\gamma_4} \\ a_{\beta_4\gamma_3} & a_{\beta_4\gamma_4} & b_{\beta_4\gamma_3} & b_{\beta_4\gamma_4} \end{array}\right| ,$$

[4] *Determinantal expression of multiple cross and dot products of polyadics in three dimensions*, by Lepine Hall Rice, this Journal, Vol. IV, No. 2.

where every 3-way minor of order 2 lying in the layers of $\mathbf{e}$'s is to be taken with its complementary (in the ordinary sense) minor lying in the layers of coefficients, the sign prefixed being the sign of the term of the determinant of (27) whose elements are the elements of the main diagonals of the two minors. Expanding each of these minors in the manner here indicated, *viz.*,

$$\begin{vmatrix} 111 & 112 \\ 121 & 122 \end{vmatrix} \begin{vmatrix} 211 & 212 \\ 221 & 222 \end{vmatrix} = 111.222 - 112.221 - 121.212 + 122.211, \quad (24)$$

and executing in each term of the $\mathbf{e}$-minor the double indeterminate multiplication called for, we have

$$\sum (\pm_{\beta_1\beta_2\beta_3\beta_4}) (\pm_{\gamma_1\gamma_2\gamma_3\gamma_4})$$

$$(\mathbf{e}_{\beta_1}\mathbf{e}_{\beta_3}\mathbf{e}_{\gamma_1}\mathbf{e}_{\gamma_2} - \mathbf{e}_{\beta_1}\mathbf{e}_{\beta_2}\mathbf{e}_{\gamma_2}\mathbf{e}_{\gamma_1} - \mathbf{e}_{\beta_2}\mathbf{e}_{\beta_1}\mathbf{e}_{\gamma_1}\mathbf{e}_{\gamma_2} + \mathbf{e}_{\beta_2}\mathbf{e}_{\beta_1}\mathbf{e}_{\gamma_2}\mathbf{e}_{\gamma_1})$$

$$(a_{\beta_3\gamma_3}\, b_{\beta_4\gamma_4} - a_{\beta_3\gamma_4}\, b_{\beta_4\gamma_3} - a_{\beta_4\gamma_3}\, b_{\beta_3\gamma_4} + a_{\beta_4\gamma_4}\, b_{\beta_3\gamma_3}); \quad (30)$$

and this evidently agrees with (19).

As another example, the multiple complement of a triadic in two dimensions $\mathbf{A}_3^{(2)} \equiv \sum \mathbf{e}_{j_1}\mathbf{e}_{j_2}\mathbf{e}_{j_3}\, a_{j_1 j_2 j_3}\ (j_r = 1, 2)$ is

$$((\mathbf{A}_3^{(2)}))_F^{(2)} \equiv \left|\begin{array}{cc|cc} \mathbf{e}_1\mathbf{e}_1\mathbf{e}_1 & \mathbf{e}_1\mathbf{e}_1\mathbf{e}_2 & \mathbf{e}_2\mathbf{e}_1\mathbf{e}_1 & \mathbf{e}_2\mathbf{e}_1\mathbf{e}_2 \\ \mathbf{e}_1\mathbf{e}_2\mathbf{e}_1 & \mathbf{e}_1\mathbf{e}_2\mathbf{e}_2 & \mathbf{e}_2\mathbf{e}_2\mathbf{e}_1 & \mathbf{e}_2\mathbf{e}_2\mathbf{e}_2 \\ \hline a_{111} & a_{112} & a_{211} & a_{212} \\ a_{121} & a_{122} & a_{221} & a_{222} \end{array}\right| \quad (31)$$

$$\equiv \sum (\pm_{\beta_1\beta_2})(\pm_{\gamma_1\gamma_2})(\pm_{\delta_1\delta_2})\mathbf{e}_{\beta_1}\mathbf{e}_{\gamma_1}\mathbf{e}_{\delta_1}\, a_{\beta_2\gamma_2\delta_2} \quad (32)$$

$$\equiv \mathbf{e}_1\mathbf{e}_1\mathbf{e}_1 a_{222} - \mathbf{e}_1\mathbf{e}_1\mathbf{e}_2 a_{221} \pm \ \ldots - \mathbf{e}_2\mathbf{e}_2\mathbf{e}_2 a_{111}. \quad (33)$$

NOTE ON DE SITTER'S UNIVERSE

By G. LEMAITRE[1]

The equations of the element of interval of a four-dimensional universe of constant positive curvature have been given by de Sitter in the form

$$ds^2 = R^2[-d\chi^2 - \sin^2\chi(d\theta^2 + \sin^2\theta d\phi^2) + \cos^2\chi d\tau^2], \qquad (1)$$

where R is a constant called the radius of the four-dimensional universe and χ, θ, ϕ, τ are coördinates. When the division of time and space is made as suggested by these coördinates, the space is itself of constant curvature and has the same radius as the universe.

In the four-dimensional universe, every point is equivalent to any other and the choice of any origin is immaterial. In the above separation of space and time, this symmetry is broken. The lines of constant coördinates χ, θ, ϕ, which are considered as the direction of the time, are not geodesics with the exception of the one $\chi = 0$, which passes through the origin. This point has now properties different from any other; it is a *center*. Free points or rays of light describe geodesics of the universe and do not move along geodesics of the space, with the exception of those which pass through the center.

It is clear that such an introduction of an apparent center in a universe which, by definition, has none is objectionable for a study of the properties of this universe. The purpose of this note is to look for a separation of space and time which is free from this objection. We shall be led to a homogenous field, non-statical and euclidean.

A homogenous division of space and time is immediately found by writing

$$ds^2 = R^2\left\{-\cosh^2\tau'[d\chi'^2 + \sin^2\chi'(d\theta'^2 + \sin^2\theta' d\phi'^2) + d\tau'^2\right\} \qquad (2)$$

With the division of time and space suggested by these coördinates, the radius of space is constant at any place, but is variable with time. It is proportional to $\cosh\tau'$.

[1]Fellow C. R. B. Educational Foundation, University of Louvain, at the Massachusetts Institute of Technology, Dept. of Physics.

The fact that the field is no more static may be urged as an objection against the significance of de Sitter's solution, but not against the way we have divided space and time to interpret it. There is, however, a serious objection against this procedure: The origin of time becomes a time absolutely distinct from every other. At the instant $\tau' = 0$, all the geodesics of the space are geodesics of the universe; at any other instant τ', there is no geodesic of the space which is a geodesic of the universe. The central point in space is removed, but now a central time has been introduced.

We shall see at once that the interval of the field may be written

$$ds^2 = R^2 \frac{-dx^2 - dy^2 - dz^2 + dt^2}{t^2} \tag{3}$$

When we separate time from space as suggested by these coördinates, *i.e.*, along the lines of constant x, y, z, we see that the geometry is euclidean. The time is now proportional to $T = \pm \int dt/t = \pm ln\ t$, and $t = 0$ represents the infinite past or future. For $t = \infty$ the distances of the points of constant dx, dy, dz tend to zero; the geodesics are asymptotic, *i.e.*, parallel, to the future or from the past. The equation of the field may also be written

$$ds^2 = R^2[-e^{-2T}(dx^2 + dy^2 + dz^2) + dT^2] \tag{4}$$

for geodesics parallel to the future, or

$$ds^2 = R^2[-e^{2T}(dx^2 + dy^2 + dz^2) + dT^2] \tag{5}$$

if parallel in the past direction.

This choice of coördinates is free from the objection of introducing a spurious assymmetry in space and time.

Before using this solution to discuss the properties of de Sitter's universe, we shall first establish the validity of Equation (3).

Using the transformation

$$\begin{cases} r = e\tau \tan \chi \\ t = e\tau \sec \chi \end{cases} \tag{6}$$

or

$$\begin{cases} \chi = \arcsin r/t \\ \tau = 1/2\ ln(t^2 - r^2), \end{cases} \tag{7}$$

the original expression (1) of de Sitter becomes

$$\frac{1}{R^2}ds^2 = -\frac{1}{1-r^2/t^2}\,\frac{(t\,dr - r\,dt)^2}{t^4} - \frac{r^2}{t^2}(d\Theta^2 - \sin^2\Theta d\phi^2) - \left(1-\frac{r^2}{t^2}\right)\frac{(t\,dt - r\,dr)^2}{(t^2-r^2)^2}$$

or

$$ds^2 = \frac{R^2}{t^2}\left[-dr^2 - r^2\,d\Theta^2 - r^2\sin^2\Theta d\phi^2 + dt^2\right]. \tag{8}$$

The expression inside the brackets is a Galilean ds^2 written in polar coördinates. By transforming to Cartesian coördinates, we get (3).

The equation is convenient for discussion, as we can easily get a representation of de Sitter's universe in a Galilean universe of Cartesian coördinates x, y, z. The lines parallel to the t-axis (x, y, z=Const.) are obviously geodesics. The rays of light are represented by the rays of light of the Galilean field (straight lines at 45° to the plane $t=0$).

The other geodesics are easily found. If $\dot{x}$, $\dot{y}$, $\dot{z}$, are the derivatives with respect to t, we must have

$$\delta\int\sqrt{1-\dot{x}^2-\dot{y}^2-\dot{z}^2}\;\frac{dt}{t}=0.$$

Euler's equations are

$$\frac{d}{dt}\left(\frac{\dot{x}}{t\sqrt{1-\dot{x}^2-\dot{y}^2-\dot{z}^2}}\right)=0$$

and similarly in y and z. Integrating,

$$\frac{\dot{x}}{\alpha t}=\frac{\dot{y}}{\beta t}=\frac{\dot{z}}{\gamma t}=\sqrt{1-\dot{x}^2-\dot{y}^2-\dot{z}^2}=\frac{1}{\sqrt{1+(\alpha^2+\beta^2+\gamma^2)t^2}}$$

where α, β, γ are integration constants, real for time-like lines, imaginary for space-like ones. Writing $C=\alpha^2+\beta^2+\gamma^2$ and integrating

$$\frac{x-a}{\alpha}=\frac{y-b}{\beta}=\frac{z-c}{\gamma}=\frac{1}{C}\sqrt{1+Ct^2}$$

where a, b, c are integration constants and therefore,

$$(x-a)^2+(y-b)^2+(z-c)^2-t^2=1/C.$$

The geodesics are equilateral hyperbolae lying in planes perpendicular to the plane $t=0$ and having an axis in this plane.

Their asymptotes are rays of light.

De Sitter separated time from space along the lines of constant χ, Θ, ϕ. From (7) these lines are now represented by a pencil of straight lines passing through the point $r=0$, $t=0$. These lines are not geodesics except the central one $r=0$, $\chi=0$ and the lines $r=t$, $\chi=\pi/2$ which are rays of light.

This explains the paradoxical singularity which occurs for this value of χ.

The Doppler-Fizeau effect is easy to calculate. Let us consider two sources of light of same proper period ds. They are supposed to describe geodesics of constant, x, y, z, then

$$\frac{1}{R}ds=\frac{dt}{t}=\frac{dt_0}{t_0}.$$

The equation of the ray of light from M to M_0 is

$$t_0=t+r;$$

for two rays, we have

$$t_0'-t_0=t'-t$$

so that the periods, as observed by an observer in M_0, are dt_0 and dt. The Doppler effect will be

$$\frac{d\lambda}{\lambda}=\frac{dt}{dt_0}-1=\frac{t}{t_0}-1=\frac{-r}{t_0}.$$

If we introduce de Sitter's coördinates, taking as center the light-source M, the observer M_0 will be at a distance

$$\sin\chi=\frac{r}{t_0};$$

we shall have,

$$\frac{d\lambda}{\lambda}=-\frac{r}{t_0}=-\sin\chi. \tag{9}$$

This expression has been given by Silberstein. He supposes the light-source at the origin. The observer describes a geodesic which neither goes through the origin, nor passes at a minimum

distance from it, but is at the intermediate case. Introducing this condition in a general expression of the Doppler effect, he finds the above formula.

In the preceding reasoning, we have supposed $t=0$ in the past so that our lines of universe converge in the future. If we reverse the sign of t and suppose that our lines are parallel in the direction of the past, we find the same value with the opposite sign. These two solutions cannot be compounded without introducing an assymmetry of the kind we are just criticizing.

We may sum up the above discussion in the following way: de Sitter's coördinates introduced a spurious inhomogeneity of the field which is not simply the mathematical appearance of center of an origin of coördinates, but really attributes distinct absolute properties to a center.

We tried to remove the difficulty by introducing other coördinates and were led to a homogeneous field; but first the field is not static and secondly, the space has no curvature. The first point may probably be accepted. Eddington writes on this subject: "It is sometimes urged against de Sitter's world that it becomes non-statical as soon as any matter is inserted in it. But this property is perhaps rather in favor of de Sitter's theory than against it."[2] Our treatment evidences this non-statical character of de Sitter's world which gives a possible interpretation of the mean receding motion of spiral nebulae.

The second point, on the contrary, seems completely inadmissible. We are led back to the euclidean space and to the impossibility of filling up an infinite space with matter which cannot but be finite. De Sitter's solution has to be abandoned, not because it is non-static, but because it does not give a finite space without introducing an impossible boundary.

[2] Eddington. The Mathematical Theory of Relativity, p. 161, Cambridge, 1923.

March, 1925.

NOTE ON QUASI-ANALYTIC FUNCTIONS[1]

By Norbert Wiener

Let $f(x)$ be an even function with period 2π. Let its Fourier series be

$$f_0/2+f_1\cos x+\ \ldots\ +f_n\cos nx+\ \ldots$$

Let us write,

$$f|g=\sum_0^\infty A_k^2 f_k g_k \qquad [A_k\neq 0].$$

Let A be the class of functions $f(x)$ for which $f|f$ is bounded. We shall term A *infinitely differentiable* if

$$\frac{1}{A_k}=o(k^{-h})$$

for all positive h, and *quasi-analytic* if it is infinitely differentiable, and if whenever $f(x)$ belongs to A and

$$f(0)=f''(0)=\ \ldots\ =f^{(2n)}(0)=\ \ldots\ =0$$

then $f(x)=0$.

The following theorems are proved without difficulty.

Theorem I. *If A is infinitely differentiable and*

$$\sum_0^\infty A_k^2 a_k^2$$

converges, there is a function $a(x)$ of A such that

$$a(x)=a_0/2+a_1\cos x+\ \ldots\ +a_n\cos nx+\ \ldots$$

[1] This paper owes its inception to a course of lectures delivered by Prof. Charles de la Vallée Poussin at the Massachusetts Institute of Technology in November, 1924. These lectures have appeared in the *Bulletin de la Société Mathématique de France*, Vol LII, section 1. The use of Fourier series for the study of quasi-analytic functions is of course due to Professor de la Vallée Poussin. The use of quadratic forms for the characterization of quasi-analytic functions follows out an idea of Carleman in several recent papers in the *Comptes Rendus*, and the expansion here developed is essentially that of Carleman.

Proof. Clearly $A_k a_k$ is bounded. Hence

$$a_k = o(k^{-h}).$$

Therefore $\Sigma |a_k|$ converges, so that

$$a_0/2 + a_1 \cos x + \ldots + a_n \cos nx + \ldots$$

converges uniformly.

Theorem II. *If A is infinitely differentiable, every function of A possesses derivatives of all orders, and vice versa.*

Proof. As above, if A is infinitely differentiable, and f belongs to A, then

$$f_k = o(k^{-h}).$$

As to the converse theorem, let A not be infinitely differentiable. Form

$$\Sigma \delta_k \cos kx$$

where

$$\delta_k = \frac{1}{k^2 A_k}; \qquad [|A_k| \geq 1]$$

$$\delta_k = \frac{1}{k^2} \qquad [|A_k| < 1].$$

Clearly $\Sigma\ \delta_k \cos kx$ converges to a function belonging to A; clearly, moreover, this function cannot be differentiated more than $2+h$ times, where h is the greatest number for which

$$\frac{1}{A_k} = O(k^{-h}).$$

Theorem III. *Let A be quasi-analytic, and let $B_k = A_k/k^2$. Let B be determined in terms of B_k, in the same way in which A is determined in terms of A_k. Then B is quasi-analytic.*

Proof. The functions of A are the second derivatives of the functions of B. Given a function belonging to A, the corresponding function belonging to B is determined except for a term $h+kx$. The k is determined by the periodicity of the functions of B. The h is uniquely determined when the value of the function at the origin is known.

Corollary. *If A is quasi-analytic,*

$$\mathbf{A}(x)=\frac{1}{2A_0}+\frac{\cos x}{A_1}+\ \cdots\ +\frac{\cos nx}{A_n}+\ \cdots$$

belongs to a quasi-analytic class, and conversely if $\mathbf{A}(x)$ belongs to a quasi-analytic class, A is quasi-analytic.

Proof. As for the direct theorem, the quasi-analytic class in question is B itself. As for the inverse theorem, if $\mathbf{A}(x)$ belongs to a quasi-analytic class B, and f belongs to A, then

$$f_k=o\left(\frac{1}{A_k}\right),$$

and f belongs to B. It then results from the definition of quasi-analyticity that A is likewise quasi-analytic. It is here and subsequently understood that no $A_k=0$.

Theorem IV. *An infinitely differentiable class A is a quasi-analytic class when and only when the class of functions $\mathbf{A}^{(2n)}(x)$ is closed over $(0, \pi)$.*

Proof. By Theorem I, a necessary and sufficient condition that $f(x)$ belong to A is that

$$\phi(x)\sim A_0f_0+A_1f_1\cos x+\ \cdots\ +A_nf_n\cos nx+\ \cdots$$

be summable and of summable square. Furthermore,

$$f^{(2n)}(0)=(-1)^n[f_1+2^{2n}f_2+3^{2n}f_3+\ \cdots\ ]$$
$$=\frac{2}{\pi}\int_0^\pi\phi(x)\mathbf{A}^{(2n)}(x)dx.$$

Hence the quasi-analyticity of A means that the functions $\mathbf{A}^{(2n)}(x)$ $[n=0, 1, \ \cdots\ .]$ admit no summable function of summable square orthogonal to them all.

Corollary. *If $f(x)$ is a member of a quasi-analytic class, and no f_n vanishes, then f and its even derivatives form a closed set over $(0, \pi)$. Under any circumstances, if f is a member of a quasi-analytic class, there is a function $g(x)$ such that if*

$$\int_0^\pi f(x)\cos nx\neq 0$$

then

$$\int_0^\pi [f(x)-g(x)]\cos nx=0,$$

while g and its derivatives form a closed set over $(0, \pi)$. *Conversely, if* $f(x)$ *is an even function of period* 2π, *possessing at all points derivatives of all orders, and if with its even derivatives it forms a closed set over* $(0, \pi)$, *it is itself a member of a quasi-analytic class.*

Theorem V. *Let A be a quasi-analytic class of functions. Let* $\{\phi_k(x)\}$ *be a set of normal and orthogonal functions closed over* $(0, \pi)$. *Let*

$$\phi_k(x)\sim a_{k0}+a_{k1}\cos x+\ \ldots\ +a_{kn}\cos nx+\ \ldots\quad [k=1, 2, \ldots]$$

Put

$$\Phi_k(x)=\frac{a_{k0}}{A_0}+\frac{a_{k1}}{A_1}\cos x+\ \ldots\ +\frac{a_{kn}}{A_n}\cos nx+\ \ldots$$

Then the functions of A are precisely those which can be represented in the form

$$f(x)=a_1\Phi_1(x)+a_2\Phi_2(x)+\ \ldots\ +a_n\Phi_n(x)+\ \ldots,$$

where $\sum_1^\infty a_n^2$ *converges. We shall have*

$$a_n=f\,|\,\Phi_n.$$

Proof. Let

$$f_n(x)=a_1\Phi_1(x)+\ \ldots\ +a_n\Phi_n(x),$$

and let

$$g_n(x)=a_1\phi_1(x)+\ \ldots\ +a_n\phi_n(x).$$

We have

$$\Phi_k\,|\,\Phi_l=\frac{2}{\pi}\int_0^\pi \phi_k(x)\phi_l(x)dx=0.$$

Hence

$$(f_m-f_n)\,|\,(f_m-f_n)=\frac{2}{\pi}\int_0^\pi [g_m(x)-g_n(x)]^2dx$$

$$=\frac{2}{\pi}\sum_{n+1}^{m} a_k^2$$

Thus

$$\lim_{m,\,n\to\infty}(f_m-f_n)\,|\,(f_m-f_n)=0.$$

Consequently $$\lim_{m,n\to\infty}\int_0^{\pi}[f_m^{(k)}(x)-f_n^{(k)}(x)]^2\,dx=0.$$

From this we deduce that $f(x)$ exists, and that $f_n(x)$ tends uniformly towards $f(x)$ with all its derivatives.

If we write

$$f(x)=\Sigma b_k\cos kx,$$

we obtain formally

$$b_k=\sum_j\frac{a_j\alpha_{jk}}{A_k}.$$

The convergence of $\Sigma a_j\alpha_{jk}$ may be deduced with the aid of a form of the Schwartz inequality, and *a fortiori* the convergence of the series for b. It is easy to verify with the aid of the classical theory of orthogonal transformations that

$$\sum_k\Big(\sum_j a_j\alpha_{jk}\Big)^2$$

converges. From this it results at once that $f(x)$ belongs to A. Indeed, the classical theory of orthogonal transformations yields us at once the result that

$$\sum_k\Big(\sum_j a_j\alpha_{jk}\Big)^2$$

converges when and only when

$$\sum_k a_k^2$$

converges. Hence the functions of A are precisely those which can be represented in the form

$$\Sigma a_k\Phi_k(x),$$

where $\Sigma\, a_k^2$ converges.

In particular, we may choose for our functions $\phi_n(x)$ the set we obtain by normalizing $\mathbf{A}^{(2n)}(x)$. Let us have

$$\phi_n(x)=A_{n0}\,\mathbf{A}\,(x)+A_{n1}\,\mathbf{A}''(x)+\ \ldots\ +A_{n,n-1}\,\mathbf{A}^{(2n-2)}(x).$$

We shall then have

$$f\Big|\Phi_n=\sum_0^{n-1}(-1)^jA_{nj}f(x)\Big|\left[\frac{d^{2j}1/dx^{2j}}{2A_0^2}+\frac{\cos x}{A_1^2}+\ \ldots\ +\frac{k^{2j}\cos kx}{A_k^2}+\ \ldots\right]$$

$$=\sum_0^{n-1}(-1)^jA_{nj}\left[f_0\frac{d^{2j}1}{dx^{2j}}+f_1+2^{2j}f_2+\ \ldots\right]$$

$$=\sum_0^{n-1}A_{nj}f^{(2j)}(0).$$

It is hence necessary that

$$\sum_{n=1}^{\infty}\left\{\sum_{j=0}^{n-1} A_{nj} f^{(2j)}(0)\right\}^2$$

should converge if f is to belong to A.

Let us now start with the supposition that

$$\sum_{1}^{\infty}\left\{\sum_{j=0}^{n-1} A_{nj}\alpha_j\right\}^2$$

converges. Form

$$f(x)=\sum_{n=1}^{\infty}\Phi_n(x)\sum_{j=0}^{n-1}A_{nj}\alpha_j$$

This will belong to A. As before, we shall have

$$f\,|\,\Phi_n=\sum_{0}^{n-1}A_{nj}f^{(2j)}(0)$$

However, by Theorem V, we have

$$f\,|\,\Phi_n=\sum_{0}^{n-1}A_{nj}\alpha_j.$$

Hence

$$\sum_{0}^{n-1}A_{nj}[f^{(2j)}(0)-\alpha_j]=0. \qquad [n=1, 2, \ldots]$$

Unless therefore there is some relation of linear dependence between the functions $\mathbf{A}_{(x)}^{(2n)}$, we must have identically

$$f^{(2j)}(0)=\alpha_j$$

Such a relation of linear dependence would mean that the function f satisfies a differential equation of finite order with constant coefficients, and hence has a terminating Fourier series. This, however, we have excluded by definition. We thus obtain:

Theorem VI. *A necessary and sufficient condition that there be a function $f(x)$ of the quasi-analytic class A such that*

$$f^{(2n)}(0)=\alpha_n \qquad [n=0, 1, 2, \ldots]$$

is that

$$\sum_{1}^{\infty}\left\{\sum_{j=0}^{n-1} A_{nj}\alpha_j\right\}^2$$

converge. The function is then

$$\sum_{n=1}^{\infty}\Phi_n(x)\sum_{j=0}^{n-1}A_{nj}\alpha_j.$$

As was pointed out to the author by Prof. de la Vallée Poussin, this theorem is at least in appearance very surprising. The A's can be so chosen as to make A consist entirely of analytic functions — for example, we may take $A_k = e^k$. In this case, any change in the value of a single a_j, excepting the first, will destroy the possibility of finding a *periodic* analytic function satisfying the assigned initial conditions. It is hence quite paradoxical to find the possibility of finding a quasi-analytic function of an assigned class with given initial values for the function and its derivatives dependent on the convergence of a series involving the initial values, and on this alone. Such, however, appears to be the case.

FILE MULTIPLICATION OF ORDERED DETERMINANTS[1]

By Lepine Hall Rice

In a previous paper[2], in which the notion of signancy was introduced, the present writer gave a rule for expressing the product of two scalar determinants of order n, of any classes p and q, and of any signancies, as a determinant of order n and of class $p+q-2$. Cayley,[3] when he first developed the theory of determinants of higher class, gave such expression to the product of two full-sign[4] determinants of any two classes not both odd. His process was a natural extension of the ordinary multiplication of two square determinants, where the result is a square determinant each of whose elements is formed by compounding a row of one factor determinant into a row of the other, the ith row into the jth

[1] See *A Theory of Ordered Determinants with Application to Polyadics*, by Frank L. Hitchcock, this journal, this volume and number, p. 205.

[2] *P-way Determinants*, Amer. Journal of Math. 40, 1918, p. 242, at p. 250.

[3] A. Cayley, *On the Theory of Determinants*, Trans. Cambridge Phil. Soc. vol. VIII (1843), p. 75; Coll. Math. Papers, vol. I. p. 63. See §2 of that paper. In the second edition of the treatise *I Determinanti*, by E. Pascal, Milano, 1923, §55, entitled Determinanti cubici e di specie superiore, after citing Günther as attributing to Vandermonde the idea of considering determinants whose elements depend on more than two indices, says: "Ma il primo che abbia tentato qualche cosa di più preciso deve reputarsi il De Gasparis . . . 1861." Making no mention of Cayley, Pascal enumerates a dozen other writers and ends with "Lecat, . . . *Abrégé sur la théorie*, etc., Gand, 1911." Turning to Lecat, we find at the beginning of his treatise a "Notice historique" of considerable length, which correctly speaks of Cayley's paper of 1843 as "un Mémoire très court mais très profond," in which Cayley "établit les propriétés fondamentales et — ceci est important à noter — *montre la nécessité de distinguer le cas ou* n [which letter Lecat uses to denote the class] *est pair de celui ou* n *est impair*. Il démontre la loi du produit de deux déterminants . . . [et] observe que la loi est en défaut si n_1 et n_2 sont tous deux impairs." At a later point, Lecat refers to Gegenbauer's work beginning in 1880 (first published in 1882), saying: "Il démontre la régle de Cayley et ramène le cas d'exception à celui où une au moins des classes est paire; d'après lui, donc, le théorème aurait lieu sans restriction, ce qui est une erreur." Turning again to Pascal we read as the last paragraph in §55: "Il Gegenbauer (op. cit.) trovò la regola di moltiplicazione dei determinanti di specie superiore. Se p, q sono le specie dei due determinanti fattori, il prodotto è di specie $p+q-2$. Cosi per $p=2$, $q=2$ si ha ancora 2 come specie del prodotto." Thus a fundamental law is attributed to Gegenbauer which was discovered and correctly stated by Cayley in 1843 and never even correctly stated by Gegenbauer.

[4] *P-way Dets.*, p. 247, a determinant "with all, or all but one, of the indices signant, according as it is of even or odd class." Not the same as though including, "persignant," a term employed by Lecat.

row constituting the ijth element of the product; or by compounding a row of the first into a column of the second; or a column into a row; or a column into a column; — four forms of product thus being possible. The n columns and n rows of a square determinant are *files* of the first and second directions, each file consisting of n elements; more generally, in a p-way determinant of order n there are n^{p-1} files of each of the p directions, each file consisting of n elements whose locants[5] are alike except for the index belonging to the direction of the file, the like part, comprising $p-1$ indices, being called the locant of the file. If then we have a p-way determinant A and a q-way determinant B and we compound the n^{p-1} files of a chosen direction in A into the n^{q-1} files of a chosen direction in B, the n^{p+q-2} compounds fall into a $(p+q-2)$-way matrix M of order n, each compound occupying the position indicated by the combined locants of the compounded files. And Cayley showed that a certain determinant of M was identically equal to the product AB, under conditions which he stated. These conditions are now merged in the rule referred to above. By this rule, the chosen directions in A and B must be signant directions, A and B being of any classes and any signancies, and the signancy of the indices in A and of those in B must follow these indices when they combine to form the indices of the product. Written as a formula, wherein for convenience the nonsignant indices are placed first and the chosen directions come last, and wherein in order to condense the work the suffixes of the indices are written under the letters which ordinarily they would follow, the rule is:

$$\left| a \underset{u_1 \ldots u_{p'}}{+} \quad \underset{u_{p'+1} \ldots u_p}{\pm} \right|^{(n)}_{p} \cdot \left| b \underset{v_1 \ldots v_{q'}}{+} \quad \underset{v_{q'+1} \ldots v_q}{\pm} \right|^{(n)}_{q} \equiv$$

$$\left| \left(\sum_{\lambda}^{n} a_{u_1 \ldots u_{p-1} \lambda} \, b_{v_1 \ldots v_{q-1} \lambda} \right) \underset{u_1 \ldots u_{p'} \, v_1 \ldots v_{q'}}{+} \quad \underset{u_{p'+1} \ldots u_{p-1} \, v_{q'+1} \ldots v_{q-1}}{\pm} \right|^{(n)}_{p+q-2} \quad (1)$$

Now (1) has been proved to be an identity. The mode of proof does not here concern us but only the fact. It follows that if the left and right members of (1) be completely expanded into terms

[5] *P-way Dets.*, p. 242.

each containing n a's and n b's, and if commutativity be invoked and a like arrangement of the $2n$ factors in all terms containing the same factors be made, then some of the terms on the right will be identical with the terms on the left, while the rest of the terms on the right will cancel out. Commutativity merely permits this like arrangement. Hence if there were no commutativity, any device which would secure a like arrangement would still give an identity. And we shall now show that when we are dealing not with scalar but with ordered determinants, devices are at hand in the shape of already defined products and developments. We shall first give some simple examples and then state the theorems.

Let **A** and **B** be two ordered determinants of class 3 and order 2, in each of which the first direction is nonsignant, the ordinal direction (marked by a star) being the first direction in **A** and the second in **B**:

$$\mathbf{A} \equiv \left| \mathbf{a}^{*}_{\mu} {}^{+}_{\pm}{}^{\pm}_{\mu}{}^{\pm}_{\mu} \right|^{(2)}_{3}, \qquad \mathbf{B} \equiv \left| \mathbf{b}^{+}_{\nu} {}^{*}_{\pm}{}^{\pm}_{\nu}{}^{\pm}_{\nu} \right|^{(2)}_{3}, \tag{2}$$

that is,

$$\mathbf{A} \equiv \begin{vmatrix} a_{111} & a_{112} & a_{211} & a_{212} \\ a_{121} & a_{122} & a_{221} & a_{222} \end{vmatrix} \equiv a_{111}a_{222} - a_{112}a_{221} - a_{121}a_{212} + a_{122}a_{211}, \tag{3}$$

$$\mathbf{B} \equiv \begin{vmatrix} b_{111} & b_{112} & b_{211} & b_{212} \\ b_{121} & b_{122} & b_{221} & b_{222} \end{vmatrix} \equiv b_{111}b_{222} - b_{112}b_{221} - b_{212}b_{121} + b_{211}b_{122}; \tag{4}$$

and let us compound the files of the last direction in **A** into the files of the last direction in **B** to form the elements of a 4-way matrix

$$M \equiv \left\| (\Sigma a_{\mu\mu\lambda} b_{\nu\nu\lambda})^{2}_{\lambda}{}_{\mu\ \nu\ \mu\ \nu}{}_{1\ 1\ 2\ 2} \right\|^{(2)}_{4} \tag{5}$$

$$\equiv \left\| \begin{array}{cc|cc} (1111) & (1112) & (1211) & (1212) \\ \Sigma a_{11\lambda} b_{11\lambda} & \Sigma a_{11\lambda} b_{12\lambda} & \Sigma a_{11\lambda} b_{21\lambda} & \Sigma a_{11\lambda} b_{22\lambda} \\ (1121) & (1122) & (1221) & (1222) \\ \Sigma a_{12\lambda} b_{11\lambda} & \Sigma a_{12\lambda} b_{12\lambda} & \Sigma a_{12\lambda} b_{21\lambda} & \Sigma a_{12\lambda} b_{22\lambda} \\ \hline (2111) & (2112) & (2211) & (2212) \\ \Sigma a_{21\lambda} b_{11\lambda} & \Sigma a_{21\lambda} b_{12\lambda} & \Sigma a_{21\lambda} b_{21\lambda} & \Sigma a_{21\lambda} b_{22\lambda} \\ (2121) & (2122) & (2221) & (2222) \\ \Sigma a_{22\lambda} b_{11\lambda} & \Sigma a_{22\lambda} b_{12\lambda} & \Sigma a_{22\lambda} b_{21\lambda} & \Sigma a_{22\lambda} b_{22\lambda} \end{array} \right\| \tag{6}$$

Then let us form that multiple determinant $\overset{|||}{\mathbf{C}}$ of the matrix M whose first two indices are nonsignant and whose last two are signant (thus continuing the signancies of **A** and **B**), and whose ordinal direction for the **a**'s is the first and for the **b**'s the last (thus continuing the ordinalities of **A** and **B**); *i.e.*, form

$$\overset{|||}{\mathbf{C}} \equiv \overset{|||}{\left| \left(\sum_{\lambda} \mathbf{a}_{\substack{u\,u\,\lambda \\ 1\,2}} \mathbf{b}_{\substack{r\,r\,\lambda \\ 1\,2}} \right)^{\substack{\mathbf{a} \\ u \\ 1}} {}_{+}^{+} {}_{r}^{\pm} {}_{1}^{\pm} \, {}_{u}^{\mathbf{b}} {}_{2}^{\pm} \, {}_{r}^{\pm} {}_{2}^{\pm} \right|}_{4}^{(2)}, \tag{7}$$

where the symbol ||| over the first bar indicates the multiple character of the determinant, that is, indicates that the two **a**'s in each term are to come first, followed by the two **b**'s; and where the superfixed **a** and **b** show the ordinality.[6] We may say that $\overset{|||}{\mathbf{C}}$ has 8 **c**-terms and 32 **ab**-terms. (In general the product has $(n!)^{p+q-3}$ **c**-terms, and $n^n\,(n!)^{p+q-3}$ **ab**-terms.) The multipleity and ordinality specified above hold of course in every **ab**-term. Signancy gives the same sign to every one of the **ab**-terms in the same **c**-term. For instance,

$$\begin{aligned} &-\mathbf{c}_{1112}\mathbf{c}_{2221} \\ &\equiv -\mathbf{a}_{111}\mathbf{a}_{221}\mathbf{b}_{211}\mathbf{b}_{121} - \mathbf{a}_{111}\mathbf{a}_{222}\mathbf{b}_{212}\mathbf{b}_{121} - \mathbf{a}_{112}\mathbf{a}_{221}\mathbf{b}_{211}\mathbf{b}_{122} - \mathbf{a}_{112}\mathbf{a}_{222}\mathbf{b}_{212}\mathbf{b}_{122}. \end{aligned} \tag{8}$$

Now let the *direct* indeterminate product **AB** be formed, *i.e.*, let the right-hand polynomial in (3) be followed by the right-hand polynomial in (4) as an indeterminate product, which by the distributive law may be written as the sum of sixteen terms, in each of which the two **a**'s come first, ordered on their first index, and the two **b**'s come last, ordered on their second index. And we see that the **ab**-terms of **AB** are identical with the same number of **ab**-terms of $\overset{|||}{\mathbf{C}}$, while the remaining terms of $\overset{|||}{\mathbf{C}}$ cancel out, so that

$$\mathbf{AB} - \overset{|||}{\mathbf{C}}. \tag{9}$$

Again, starting with the same **A**, **B**, and M, let the *multiple* indeterminate product $\mathbf{A} \,|||\, \mathbf{B}$ be formed, so that in each term the **a**'s and **b**'s alternate, and let the *direct* determinant **C** of the

[6] In a multiple ordered determinant the ordering is effected in each term, after the term has been formed by taking a transversal of the matrix, prefixing a sign in accordance with the signancy, and grouping the factors in accordance with the multipleity.

matrix M be formed, in signancy and ordinality precisely like $\overset{|||}{\mathbf{C}}$. In the **ab**-terms of **C** the **a**'s and **b**'s will alternate, and therefore

$$\mathbf{A} \,|||\, \mathbf{B} \equiv \mathbf{C}. \tag{10}$$

Evidently the first case illustrates the theorem:

Theorem 1. *If* **A** *and* **B** *are ordered determinants of the matrices*

$$\left\| \mathbf{a}_{u_1 \dots u_p} \right\|_p^{(n)}, \quad \left\| \mathbf{b}_{v_1 \dots v_q} \right\|_q^{(n)} \tag{11}$$

of any signancies and ordinalities so long as u_p *and* v_q *are signant and not ordinal, and* **AB** *denotes the direct product, and if* $\overset{|||}{\mathbf{C}}$ *is that multiple determinant of the matrix*

$$M \equiv \left\| \left(\sum_{\lambda}^{n} \mathbf{a}_{\substack{u \dots u \lambda \\ 1 \quad p-1}} \mathbf{b}_{\substack{v \dots v \lambda \\ 1 \quad q-1}} \right)_{\substack{u \dots u \; v \dots v \\ 1 \quad p-1 \; 1 \quad q-1}} \right\|_{p+q-2}^{(n)} \tag{12}$$

in which the signancies and ordinalities of the u's and v's are continued, then

$$\mathbf{AB} \equiv \overset{|||}{\mathbf{C}}. \tag{13}$$

The second case similarly illustrates the theorem:

Theorem 2. *If* **A**, **B**, *and* M *are as in* Theorem 1, *and* $\mathbf{A} \,|||\, \mathbf{B}$ *denotes the multiple product, and if* **C** *is that direct determinant of* M *in which the signancies and ordinalities of the u's and v's are continued, then*

$$\mathbf{A} \,|||\, \mathbf{B} \equiv \mathbf{C}. \tag{14}$$

A THEORY OF ORDERED DETERMINANTS WITH APPLICATION TO POLYADICS

By

FRANK L. HITCHCOCK

Table of Contents

Introduction

Part I. Vector Determinants and their Element Multiplication

1. Laws of Expansion
2. Types of Multiplication
3. Extension of the Concept of Ordinal Directions
4. A Multiplication Theorem
5. The Scott-Rice Element Multiplication
6. Dummy Indices
7. Application to Scalars
8. Conjugate Theorems

Part II. Application to Polyadics

9. Definition and Elementary Properties of Polyadics
10. Determinants Obtained by Folding k-adic Expansions
11. The Scalar Components of the Polyadics F_k
12. Folding a Polyadic into Itself

Introduction. The stimulus toward the present investigation is to be found in a paper by my colleague Mr. Lepine Hall Rice (This Journal Vol. IV, No. 2, Page 130 (April 1925), *Determinantal Expression of Multiple Products*, etc.) in which it is shown that multiple cross products of polyadics in three -space can be expressed as determinants of class greater by one than the degree of the polyadics. Some of the elements of such determinants are vectors, dyads, or in general polyads.

In extending this result to space of n dimensions and to all polyadics we have been led to study determinants whose elements

are not commutative in multiplication. The polyadic is the most general element of this kind; but the laws which must be obeyed in the development of the determinant are for the most part independent of the specific application to be made of the result. The first part of the present paper is devoted to questions of ordered development, made necessary by the non-commutative nature of the elements. The Scott-Rice " element multiplication " of determinants is shown to hold good and several theorems are obtained on determinants with polyad elements. In a paper, p. 200, this number, Mr. Rice shows that " file multiplication " also holds for determinants with non-commutative elements.

The applications of these determinants are numerous. As already remarked, the polyadic is the most general non-commutative element. It is therefore natural, as a first application, to deal with p-way determinants which are functions of one or of several polyadics. In a joint paper Mr. Rice and the author have shown (This Journal, Vol IV, No. 3, Page 179) that by a process called *folding* a generalization of multiple cross product is obtained, the result being expressible in determinant form. In this process the " space complement " of Dr. Almar Naess is fundamental. In Part II of the present paper is a general treatment of the folding process. It is shown that from a given set of polyadics of degree p, families of invariants may be obtained which fall easily into series $F_0, F_1, \ldots F_{p-1}$ according to the number of folded factors of the polyadics. A few of the simpler properties of these quantities are examined, and a formulation of the general scalar component of F_k is given.

A detailed application of these various sets of invariants is not attempted in this paper. It may be remarked, however, that it is the series F_o which constitutes the generalization of the multiple cross product above mentioned. And just as in the well-known case of dyadics, the vanishing of double cross " powers " of different degrees determines the degree of nullity of the dyadic, so, in general, the question of the minimum number of terms by means of which a polyadic may be expressed undoubtedly finds its answer in the vanishing of one or several of the quantities F_k.

Again, it is well known that to every polyadic corresponds a multilinear " form," and to every symmetrical polyadic corre-

sponds a polynomial. In the same way the quantities F_k correspond to rational, integral, concomitants of such forms. For example, transvection is equivalent to the process F_{p-1}, performed on n polyadics. The folding process is closely analogous to the usual symbolic method of obtaining these concomitants, but differs from the symbolic method in being *actual*, in a numerical sense, at every step, *i.e.*, if the given polyadics have assigned numerical components. the same is true of F_k in all its transformations. Whether this fact is, or is not, a great advantage, must be perhaps a matter of individual judgment. The " weight " of an invariant is the same as the number of " folds."

Another fairly obvious application is to operational methods. The well-known equation in dyadics $\phi' \cdot \overset{\times}{\phi}_{\times} \phi = mI$ is substantially a case of the operation F_1.

It has been suggested that these p-way non-commutative combinations may be useful in problems of analysis situs, but the connection, if any, has not yet been found.

Since the basis vectors as employed in the present paper may be regarded as defining an affine space, results are directly applicable in tensor calculus in so far as the properties of such a space are concerned in the problem. For as is well known a tensor possesses a formal resemblance to a polyadic which in many transformations amounts to complete identity.

With regard to the difficult question of nomenclature, the following terms appear to be in agreement with the usage of others.

1°. The *dimensions* n of the space considered is the number of basis vectors $\mathbf{e}_1, \mathbf{e}_2, \ldots \mathbf{e}_n$.

2°. The *class* of a determinant is the number of *ways* or *directions* which it possesses, equal, usually, to the number of indices or other device appended or belonging to each element.

3°. The *order* of a determinant is the number of elements in each *file* or row or column. This number is frequently, but not necessarily, the same as n.

4°. The *degree* of a polyadic is the number of *factors* in each trem.

In this paper the number of factors not folded is denoted by k. The term *folding*, while not free from objection, is suggested

by the fact that, in the process, a group of factors is made into a new unit, denoted by a single letter **E**. The number of folded polyadics is denoted by q. Thus Gibbs' double cross product of two dyadics ϕ_1 and ϕ_2 falls into the scheme as $\mathbf{F}_0((\phi_1, \phi_2))^3$, where the upper index denotes the dimensions of the space. It need scarcely be said that the above nomenclature is tentative, and is likely to require modification in specific applications.

The "general term" of any F_k is given by the expressions (38) and (39) of Art. 11. It consists of a polyadic made up of fundamental units, multiplied by a scalar determinant of class $p-k+1$. It contains three sets of indices, of which two sets are denoted by italics and have fixed values in any chosen term. The third set is denoted by small capitals and gives the class of the scalar coefficient. The utility of ordered determinants is exemplified in the ease by which any required term can be picked out, on the one hand, and, on the other, general properties of F_k may be proved. Without the light afforded by such determinants the general term of F_k would doubtless appear too complex to be practicably handled.

Part I. Vector Determinants and Their Element Multiplication

1. Laws of Expansion.

When the elements of a *two-way* determinant are not commutative in multiplication it is usual to follow Cayley's rule that the elements shall stand in the order of the rows, thus

$$\begin{vmatrix} \alpha_1, & \alpha_2, & \alpha_3 \\ \beta_1, & \beta_2, & \beta_3 \\ \gamma_1, & \gamma_2, & \gamma_3 \end{vmatrix} = \alpha_1\beta_2\gamma_3 - \alpha_1\beta_3\gamma_2 + \alpha_2\beta_3\gamma_1 - \alpha_2\beta_1\gamma_3 + \alpha_3\beta_1\gamma_2 - \alpha_3\beta_2\gamma_1, \tag{1}$$

and similarly for square determinants of any order. While I shall speak of the elements of such determinants as vectors, they may, in much of what follows, be any quantities or operations not in general commutative, such as quaternions, polyadics, or differential or integral operators.

Determinants whose elements are actually vectors are closely related to scalar determinants of higher class. For example, let

the elements of (1) be vectors in space of three dimensions defined by the equations

$$\alpha_i = \mathbf{e}_m a_{im}, \quad \beta_j = \mathbf{e}_m b_{jm}, \quad \gamma_k = \mathbf{e}_m c_{km}, \text{ (summed for } m = 1, 2, 3) \tag{2}$$

where $\mathbf{e}_1$, $\mathbf{e}_2$, and $\mathbf{e}_3$ are basis vectors; then from each term on the right of (1) set up the scalar two-way determinant of the components of the vectors which occur in it, thus

$$(\alpha_1\beta_2\gamma_3) = \begin{vmatrix} a_{11}, & a_{12}, & a_{13} \\ b_{21}, & b_{22}, & b_{23} \\ c_{31}, & c_{32}, & c_{33} \end{vmatrix} \tag{3}$$

and similarly for the other five terms. The algebraic sum of the six square determinants thus obtained is equal to the cubic determinant

$$\begin{vmatrix} a_{11}, & a_{12}, & a_{13} \\ a_{21}, & a_{22}, & a_{23} \\ a_{31}, & a_{32}, & a_{33} \end{vmatrix} \begin{vmatrix} b_{11}, & b_{12}, & b_{13} \\ b_{21}, & b_{22}, & b_{23} \\ b_{31}, & b_{32}, & b_{33} \end{vmatrix} \begin{vmatrix} c_{11}, & c_{12}, & c_{13} \\ c_{21}, & c_{22}, & c_{23} \\ c_{31}, & c_{32}, & c_{33} \end{vmatrix} \tag{4}$$

where the direction from a to b to c is non-signant. (The definitions of p-way scalar determinants and of signant directions which I shall employ are those of Lepine Hall Rice, *American Journal of Mathematics*, Vol. XL, No. 3, July 1918, a paper fundamental to the subject, which I shall refer to as *Rice*). This relation between (1) and (4) is a case of a more general theorem to be proved below. The importance and interest attaching to vector determinants is however not merely or even chiefly the possibility which they afford of expressing scalar p-way determinants by means of a smaller number of formal directions; it resides rather in the invariant properties which these determinants possess and in the combinatory operations to which they give rise, applicable to a great variety of non-commutative elements.

Determinants of more than two directions whose elements are not commutative in multiplication have not before been studied so far as I know. It is therefore necessary to establish rules for their expansion. We assume n^p elements. These will be distinguished by p indices or other devise said to correspond to *directions*. For the present it is assumed that each element consists of a single vector. Let one index or direction be distinguished from the others and named *ordinal*, and, in every term of the

expansion let the order of elements be their order according to this direction. In (1), the ordinal direction is the vertical direction, distinguished by alphabetical order of letters instead of by an index, a devise often convenient but not essential. The meaning of the vector determinant is then fully determined (by *Rice*, page 246). The simplest examples are cubic determinants of the second order. If we make the ordinal direction the same as the non-signant direction we shall have

$$\begin{vmatrix} \alpha_{11}, & \alpha_{12} \\ \alpha_{21}, & \alpha_{22} \end{vmatrix} \begin{vmatrix} \beta_{11}, & \beta_{12} \\ \beta_{21}, & \beta_{22} \end{vmatrix} = \alpha_{11}\beta_{22} - \alpha_{21}\beta_{12} - \alpha_{12}\beta_{21} + \alpha_{22}\beta_{11} \quad (5)$$

where both subscripts are signant and where the direction from α to β is non-signant as well as ordinal. Thus if each α were to be interchanged with the corresponding β, the determinant would be unaltered if its elements were commutative but is here altered because $\alpha\beta$ is not the same as $\beta\alpha$.

On the other hand if the ordinal direction is the same as one of the signant directions we shall have

$$\begin{vmatrix} \alpha_{11}, & \alpha_{12} \\ \beta_{11}, & \beta_{12} \end{vmatrix} \begin{vmatrix} \alpha_{21}, & \alpha_{22} \\ \beta_{21}, & \beta_{22} \end{vmatrix} = \alpha_{11}\beta_{22} - \alpha_{22}\beta_{11} - \alpha_{12}\beta_{21} + \alpha_{21}\beta_{12} \quad (6)$$

where the first subscript is non-signant and the ordinal direction is still from α to β, meaning that, in the expansion, every α precedes every β.

Vector determinants possess, in common with scalar determinants, the property that two or more non-signant layers may be alike, and have also the new property that two or more ordinal layers may be alike, without in general entailing the vanishing of the determinant. Thus neither (5) nor (6) becomes identically zero if every α is equal to the corresponding β. In (6), where the first index is non-signant, we may not only make each α equal to the corresponding β but may at the same time suppress the first subscript, when the determinant reduces to $2(\alpha_1\alpha_2 - \alpha_2\alpha_1)$. On the other hand (5) vanishes identically if we suppress the first subscript because this index is signant but not ordinal.

Similar principles apply to any number of directions, and for determinants of any order. The expansion is determined by two new concepts,— new, that is, as compared with ordinary two-way

scalar determinants, — first, Rice's rule of signs, and second, the ordinal direction. It is to be emphasized that *the rule of signs is quite independent of the ordinal direction*, and is to be applied precisely as if the elements were commutative. This fact is important, for in consequence, as will be shown, many theorems known to be true for scalars can be at once extended to vectors or other non-commutative elements by attention to matters of order.

To illustrate, the term $\alpha_{21}\beta_{12}$ is negative in (5) because, both indices being signant, the second index runs in the natural order while the first has suffered a permutation. This term is positive in (6) because the first index is non-signant, the second index is in the natural order, and the order α, β is the natural order. By Rice's rule the sign is minus or plus according as the number of permutations among signant indices (or equivalent) is odd or even.

2. Types of multiplication.

Before proceeding to the proof of the fact that the product of two or more vector determinants is a determinant it will be necessary to consider determinants whose elements are not single vectors but dyads in the sense of Gibbs. In general we may say that a dyad element consists of the product of two non-commutative factors. These are merely written one after the other with no sign between. For example, we may write the four-way determinant of the second order

$$\left|\begin{array}{cc|cc} \alpha_{11}\beta_{11}, & \alpha_{11}\beta_{12} & \alpha_{12}\beta_{11}, & \alpha_{12}\beta_{12} \\ \alpha_{11}\beta_{21}, & \alpha_{11}\beta_{22} & \alpha_{12}\beta_{21}, & \alpha_{12}\beta_{22} \\ \hline \alpha_{21}\beta_{11}, & \alpha_{21}\beta_{12} & \alpha_{22}\beta_{11}, & \alpha_{22}\beta_{12} \\ \alpha_{21}\beta_{21} & \alpha_{21}\beta_{22} & \alpha_{22}\beta_{21}, & \alpha_{22}\beta_{22} \end{array}\right| \tag{7}$$

and it will be shown that, with proper rules as to ordinal directions, etc., this determinant is equal to the product of two simpler determinants. More generally we may have determinants whose elements are polyads or polyadics.

In multiplying polyads or polyadics together many types of product are possible, but two are of especial importance. Take

for example two dyads $\alpha_1\beta_1$ and $\alpha_2\beta_2$. We may form what I shall call their *direct product* by writing them one after the other with no sign between as $\alpha_1\beta_1\alpha_2\beta_2$. Or we may form a "double" product in the Gibbs manner by stipulating that the α's shall precede the β's. This "double indeterminate product" will be $\alpha_1\alpha_2\beta_1\beta_2$.

More generally, if we have several polyads $\alpha_1\beta_1\gamma_1 \ldots \eta_1$ and $\alpha_2\beta_2\gamma_2 \ldots \eta_2$ and so on, the respective α's are called the *first factors* of their polyads, the β's are *second factors*, and so on.

Definition I. The *direct product* of several polyads is formed by writing them one after the other with no sign between.

Definition II. The *multiple indeterminate product* of several polyads all of which have the same number of factors is formed by writing all the first factors, then all the second factors, and so on. The order among the factors of each group is the same as the order of multiplication of the polyads.

The multiple indeterminate product may be abbreviated m. i. p. It is obvious that these two general types of multiplication may be specialized by the introduction of dot, cross, or alternant product in various ways, or by replacing the vectors, some or all, by various operational symbols, for these two definitions have to do only with the order of the quantities concerned. Both types may be extended to polyadics as usual.

3. **Extension of the concept of ordinal directions.**

If the elements of a determinant are polyads of the same number of factors as in (7), a development of the determinant such that, in every term, all first factors precede all second factors, and so on, will be called an *m. i. p. expansion*.

If we decide that the expansion shall be by m. i. p. the order of the first factors among themselves is still undetermined, as is also the order of the second factors among themselves, and so on. We therefore choose an ordinal direction or index which shall determine the order of the first factors among themselves. This will, of course, be an index or other devise which belongs to the first factors. We then choose one of the directions pertaining to the second factor to be the ordinal direction for the second factor,

and so on. When we have also chosen the non-signant directions (if any), the development is fully determined. For example, in (7), we may first decide that all directions shall be signant. The four directions within the determinant are determined by two indices pertaining to the α's and two for the β's. If we decide that the first index shall be the ordinal index in each case the development is unique provided all α's precede all β's, that is, if the expansion is by m. i. p. The expansion then consists of eight terms, namely

$$\begin{aligned}&a_{11}a_{22}\beta_{11}\beta_{22}-a_{11}a_{22}\beta_{12}\beta_{21}-a_{11}a_{22}\beta_{12}\beta_{21}+a_{11}a_{22}\beta_{11}\beta_{22}\\&-a_{12}a_{21}\beta_{11}\beta_{22}+a_{12}a_{21}\beta_{12}\beta_{21}+a_{12}a_{21}\beta_{12}\beta_{21}-a_{12}a_{21}\beta_{11}\beta_{22}\\&=2(a_{11}a_{22}-a_{12}a_{21})(\beta_{11}\beta_{22}-\beta_{12}\beta_{21})\end{aligned}\tag{7'}$$

where, in the last line, the multiplication is direct instead of m. i. p.

4. A multiplication theorem.

In (7) and (7') we have an example of a four-way determinant which, *when developed by m. i. p.*, is equal to the *direct* product of two two-way determinants, together with the scalar factor 2.

A general theorem may be stated as follows:

Theorem I. A determinant of class $p+q$ whose elements are of the form

$$a_{i_1\ldots i_p}\beta_{j_1\ldots j_q}\tag{8}$$

is the *direct product* of the determinants of respective classes p and q whose elements are of the form $a_{i_1\ldots i_p}$ and $\beta_{j_1\ldots j_q}$, multiplied by $n!$; provided the following rules are observed in the expansion of the determinant of class $p+q$:

1°. The development is by m. i. p., that is, all α's precede all β's.

2°. The ordinal indices are the same as in the factor determinants.

3°. The indices are signant or not according as they are signant or not in the factor determinants.

Proof. Rules 1° and 2° insure the correct order among the factors of the individual terms. Rule 3° insures, by Rice's defini-

tion of signancy, that the sign of any term of the expansion is correct. And any chosen term occurs $n!$ times. Hence the theorem holds.

As particular cases, all the α's may be scalars, or all the β's may be scalars, or both. If all elements are scalars the distinction as to ordinal indices and m. i. p. becomes unnecessary. If the α's are scalars while the β's are vectors there will remain an ordinal index for the β's, and vice versa.

Extension to any number of determinants is evident and may be stated as follows:

Theorem Ia. If every element of a determinant of class $p_1+p_2+\ldots+p_k$ can be written as the indeterminate product of k factors carrying respectively $p_1, p_2, \ldots, p_k$ indices, any *m. i. p. expansion* is equal to $(n!)^{k-1}$ times the *direct product*, in the order of the factors, of k determinants of respective classes $p_1, p_2, \ldots,$ and p_k whose elements are the corresponding factors of the first determinant; provided signant and ordinal character of each index is continued throughout.

The requirement that the signancy of every index shall be continued shows, on the one hand, that given determinants can always be multiplied under this theorem; but that on the other hand a given determinant with k-ad elements can be factored (under this rule at least) only when the signancy of its indices is such that the required determinants of lower class exist. For example a four-way determinant with elements $\alpha_{ij}\beta_{rs}$ can be factored if all indices are signant, as in (7) and (7′). It can also be factored if i and j are both non-signant, or if r and s are both non-signant, or if all four indices are non-signant, for in these cases one (or both) of the two-way factor determinants merely becomes a two-way determinant with both directions non-signant, that is, a permanent (Rice, page 246). But if, say, i and r are signant while j and s are non-signant, the conditions of theorem I cannot be fulfilled. It is easy to verify by direct expansion that, under this last hypothesis, any m. i. p. expansion of (7) vanishes identically. Why it must do so will presently appear.

5. **The Scott-Rice element multiplication.**

The element multiplication of scalar determinants (Rice, page 249) while formally a little less simple than the preceding rule, is actually more so, since fewer directions are involved and the scalar factor $n!$ does not enter. Rice's prescription extends at once to determinants whose elements are k-ads if we attend to the ordinal indices and if the development is by m. i. p. The result may be stated thus:

Theorem II. If every element of a determinant of class $p+q-1$ can be written as the indeterminate product of two factors carrying respectively p and q indices, wherein a pair of indices, one from each factor, are regarded as the same index, and determine but one direction, any m. i. p. expansion is equal to the direct product, in the order of the factors, of two determinants of respective classes p and q whose elements are the corresponding factors of the first determinant; provided that

1°. The paired index is signant in the determinant with dyad elements only when it is signant in one of the simple determinants but not in the other.

2°. The signancy of the other indices is continued.

3°. The ordinal character of all indices is continued. This does not imply that the paired index is paired as to ordinal character: it may be ordinal for one factor but not for the other.

These rules obviously insure the correct order of all factors, and the signancy is in agreement with Rice's rule.

As in the former case, we can always multiply under the theorem, and it will be shown later that we can always factor.

It is clear that we may multiply one determinant into another in a number of ways, according to the signancy and ordinal character of the paired indices. For example, the direct product of $(\alpha_{11}\alpha_{22}-\alpha_{12}\alpha_{21})$ into $(\beta_{11}\beta_{22}-\beta_{12}\beta_{21})$, where the first index is ordinal in both cases, may be written as a cubic determinant by pairing the ordinal indices, now becoming non-signant in the cubic determinant, giving

$$\begin{vmatrix} \alpha_{11}\beta_{11}, & \alpha_{11}\beta_{12} & \alpha_{21}\beta_{21}, & \alpha_{21}\beta_{22} \\ \alpha_{12}\beta_{11}, & \alpha_{12}\beta_{12} & \alpha_{22}\beta_{21}, & \alpha_{22}\beta_{22} \end{vmatrix} \tag{9}$$

as may be verified by m. i. p. expansion. On the other hand if we had paired on the two non-ordinal indices we should have had

$$\begin{vmatrix} \alpha_{11}\beta_{11}, & \alpha_{11}\beta_{21} & \alpha_{12}\beta_{12}, & \alpha_{12}\beta_{22} \\ \alpha_{21}\beta_{11}, & \alpha_{21}\beta_{21} & \alpha_{22}\beta_{12}, & \alpha_{22}\beta_{22} \end{vmatrix} \tag{10}$$

which again verifies by m. i. p. expansion , not forgetting that the first indices in each factor are still ordinal. Again, we can pair the ordinal or first index from the α's with the second or non-ordinal index from the β's, giving

$$\begin{vmatrix} \alpha_{11}\beta_{11}, & \alpha_{11}\beta_{21} & \alpha_{21}\beta_{12}, & \alpha_{21}\beta_{22} \\ \alpha_{12}\beta_{11}, & \alpha_{12}\beta_{21} & \alpha_{22}\beta_{12}, & \alpha_{22}\beta_{22} \end{vmatrix} \tag{11}$$

and finally we can pair the second index of the α's with the first of the β's,

$$\begin{vmatrix} \alpha_{11}\beta_{11}, & \alpha_{11}\beta_{12} & \alpha_{12}\beta_{21}, & \alpha_{12}\beta_{22} \\ \alpha_{21}\beta_{11}, & \alpha_{21}\beta_{12} & \alpha_{22}\beta_{21}, & \alpha_{22}\beta_{22} \end{vmatrix} \tag{12}$$

The last two examples illustrate the fact that the paired index may be ordinal for one factor without being so for the other. Thus the four determinants (9)-(12) are equal, if expanded in the manner defined under theorem II.

Taking the next simplest case, the direct product of a two-way determinant into a three-way determinant will be a four-way determinant with two non-signant directions if we pair on two signant indices, but will be of full sign if one of the paired indices is non-signant. In general all relations of signancy are the same as for scalar determinants.

6. Dummy indices.

As shown in Art. 1 vector determinants may have either non-signant or ordinal layers alike. Conversely, from a given determinant we may form another of one more direction by appending to every element a new index which shall be a *dummy* in the sense that, *as this index takes on different values from* 1 *to n, the element is actually unaltered.* The dummy is not quite fictitious, however, because the new determinant is not equal to the old. Thus if to the elements of the two-way determinant $\alpha_{11}\alpha_{22}-\alpha_{12}\alpha_{21}$ we append a non-signant dummy we obtain, in form, eight ele-

ments which develop into a cubic determinant. After the expansion the dummy may be suppressed, and the result is twice the original determinant. In general we have the following:

Lemma. A non-signant, non-ordinal dummy index may be attached to the elements of any determinant. The determinant of one more direction thus obtained is equal to the original determinant multiplied by $n!$.

It can now be seen why, as mentioned at the end of Art. 4, a determinant of the form (7) must vanish identically if there are only two signant indices and if these are chosen one from each factor. For let the elements be $\alpha_{ij}\beta_{rs}$ and let i and r alone be signant. Let a paired dummy ζ be appended to every factor, giving a determinant of five directions whose elements are $\alpha_{ij\zeta}\beta_{rs\zeta}$. The direction ζ must be non-signant for this determinant, which, by the lemma, is equal to $n!$ times the original determinant. We can also have cubic determinants whose elements are $\alpha_{ij\zeta}$ and $\beta_{rs\zeta}$ respectively in which ζ is signant, the non-signant directions being j and s. By theorem II the m. i. p. expansion of the five-way determinant is the direct product of the two cubics. But these both vanish indentically since ζ is a signant dummy and not ordinal. Hence the original determinant is zero.

Similar reasoning may be employed in any case. The result may be stated as:

Theorem III. If every element of a determinant of class $p_1+p_2+\ldots+p_k$ can be written as the indeterminate product of k factors carrying $p_1, p_2, \ldots p_k$ indices respectively, any m. i. p. expansion vanishes identically in case any factor carries an odd number of signant indices.

We note that this is precisely the case where the determinant with k-ad elements cannot be factored under theorem I. Thus such a determinant can either be factored or vanishes identically.

Proof. If any factor carries an odd number of signant indices there must be an even number of such factors, because the total number of signant indices is by definition even. Let the factors which carry an odd number of signant indices be taken in pairs and to each pair let a paired non-signant, non-ordinal dummy index be attached. The resulting determinant of higher class is,

by the lemma, equal to the given determinant aside from a scalar factor. If the dummies be made signant on the separate factors, the determinant may be factored under theorem II, by successive application of the theorem if necessary. But any factor determinant which carries a dummy vanishes identically, since the dummy is signant on the factors separately. Hence the original determinant vanishes identically.

We note also that theorem III is independent of the ordinal character of the indices, assuming only that the factors, if they are not scalars, possess an ordinal index no matter which. Evidently if an expansion vanishes identically with a given distribution of ordinal indices among the factors, it will vanish under any other distribution, since a change of ordinal indices affects all like terms in the same manner.

By contrast it may be pointed out that a determinant with paired indices, of the type contemplated under theorem II, *cannot vanish identically when no index is a dummy*, for no two terms of the expansion are like terms. Furthermore we can always factor. For suppose an impossible case, that is, assign to each factor of the dyad elements an odd number of signant indices. By reversing the signancy of both the paired indices the determinant is unchanged, but the factors now carry an even number of signant indices. Hence factoring is possible.

7. Application to scalars.

If the elements of a determinant of the n^{th} order and any class are vectors in n-space, that is if any $n+1$ are linearly related, each term in the expansion will be the indeterminate product of n vectors in a definite order. If, for every term, we form the determinant of the components of these n vectors, as illustrated in Art. 1, and add the results with proper sign, the resulting scalar may be called *the scalar of the determinant*. If the elements of the vector determinant be $\alpha_{ij\ldots}$ the scalar of the determinant may be denoted by $(\alpha_{ij\ldots})$.

Again if we have a determinant with k-ad elements as in theorems I and II, each term of an expansion by m. i. p. will be the indeterminate product of kn vectors. If in every term we form the determinant of the scalar components of the n first fac-

tors, also that of the n second factors, and so on, the algebraic sum of the products of the scalar determinants thus formed will be called the scalar of the determinant and may be denoted by $(\alpha_{ij}\ldots\beta_{rs}\ldots\ldots)$ or similar notation.

The following theorem may now be proved:

Theorem IV. If the elements of a determinant of class p are vectors each carrying p indices, its scalar can be written as a determinant of class $p+1$.

Proof. Let every element $\boldsymbol{\alpha}$ be expanded as $\Sigma\mathbf{e}_m a_m$ where m is a new index and where every $\boldsymbol{\alpha}$ and the corresponding a's carry p other indices not here written. Form a determinant of the a's with m signant. Now the $\boldsymbol{\alpha}$'s carry an ordinal index. Let its signancy be reversed for the a's. Let the signancy of the other indices be continued. The result is the scalar of the original vector determinant.

For each required term occurs once only. The signancy of m follows from the fact that, in taking the scalar of the vector determinant, we form from each term of the expansion a two-way scalar determinant one of whose directions corresponds to m. The continuation of the signancy of a non-ordinal index is implied in the definition of the scalar of a vector determinant. Finally to see that the signancy of the ordinal index must be reversed in the determinant of the a's, note that if, after expanding the vector determinant as in (1) and taking the scalar of each term, we interchange two ordinal layers, the sign of the scalar determinant formed from each term is changed; while besides this the sign of this term in the vector expansion like (1) is changed or not according as the ordinal direction is signant or not. Hence, if it is signant the sign of the resulting scalar is changed twice, *i.e.*, left unaltered. Hence the signancy of the ordinal index is reversed in the a's.

As an immediate consequence we have

Theorem V. If the elements of a determinant of p directions are vectors each carrying p indices, and if the ordinal layers are non-signant and alike, the scalar of the determinant vanishes identically.

For the scalar can be written as a determinant with signant

layers alike. The simplest example is the two-way permanent $a_1a_2+a_2a_1$ whose scalar evidently vanishes, the a's being two-space vectors. Another example of some importance is the cubic determinant with elements vectors in three-space of the form a_{ijm} where m is a non-signant ordinal dummy. The vector expansion contains 36 terms which are all different even if the dummy be suppressed after the expansion. But the scalar is zero as may be easily verified.

If we have any equation satisfied by non-commutative quantities, and if all products on both sides are indeterminate, the equation will still hold if the multiplication be specialized in the same manner on both sides. Thus if vector determinants are multiplied to form a determinant of higher class, as in theorems I and II and Ia, the scalar of the determinant of higher class is equal to the product of the scalars of the factor determinants, with or without the scalar factor $n!$ as the case may be. Whence it follows that *the scalar of a determinant with k-ad elements* (as in theorem Ia), *can be expressed as a scalar determinant of class greater by k.* For by theorem IV the scalars of the factors can be written as scalar determinants of respective classes p_1+1, p_2+1, . . . , p_k+1. Multiplying these scalar determinants by theorem Ia, (which of course applies to scalars as well as vectors) we obtain a determinant of class $p_1+p_2+\ldots+p_k+k$ each of whose elements is the product of k scalars, which carry respectively p_1+1, p_2+1, . . . , p_k+1 indices.

8. **Conjugate theorems.**

If we have any identical equation satisfied by non-commutative quantities, and if all products on both sides are indeterminate, the equation will still hold if the factors of all terms are permuted in the same way. From a proven theorem we may thus obtain other theorems which may be called conjugates of the first.

For example, theorems I, II, and Ia deal with the direct product of determinants with vector elements. If instead we form the multiple indeterminate product of the same determinants, the resulting expansion will contain identically the same terms (with the same signs) except for the order of factors in the several terms. Keeping the former notation, if a determinant with elements a

(carrying p indices) be multiplied into a determinant with elements β (carrying q indices) and if the multiplication be m. i. p. instead of direct, in the resulting expansion we shall have, instead of all α's preceding all β's an α alternating with a β throughout. The order of α's and β's considered separately is determined by the given ordinal indices as before. Thus the multiple indeterminate product of $\alpha_{11}\alpha_{22}-\alpha_{12}\alpha_{21}$ into $\beta_{11}\beta_{22}-\beta_{12}\beta_{21}$ is

$$\alpha_{11}\beta_{11}\alpha_{22}\beta_{22}-\alpha_{11}\beta_{12}\alpha_{22}\beta_{21}-\alpha_{12}\beta_{11}\alpha_{21}\beta_{22}+\alpha_{12}\beta_{12}\alpha_{21}\beta_{21} \qquad (13)$$

On the other hand if we have a four-way matrix whose elements are $\alpha_{ij}\beta_{rs}$, namely the matrix (7), we may develop it according to Rice's rule so far as signs are concerned, just as if the elements were scalars, and may further stipulate that an α shall alternate with a β throughout. We may define this as a *direct expansion*. The expansion is not yet fully determined, but becomes so if we choose ordinal indices to agree with (13), namely the first index for both α and β. The result must be the double of (13), for what we have done, in fact, is to permute both sides of (7′) in a definite way.

More generally we may define a *direct expansion* of a determinant with k-ad elements to be one where the respective factors of these elements occur in rotation. Each term in the expansion contains nk factors, and may be thought of as containing n groups, each group containing k factors. In each group we have one of the first factors of the given elements followed by one of the second factors, and so on, the group ending with one of the last or k^{th} factors. The order of first factors, second factors, etc., considered separately, must be determined by ordinal indices, assigned for each factor.

To illustrate, if we have a six-way matrix of the second order with triad elements $\alpha_{ij}\beta_{pq}\gamma_{rs}$, a direct expansion with i, p, r ordinal might begin with the term $\alpha_{11}\beta_{11}\gamma_{11}\alpha_{22}\beta_{22}\gamma_{22}$, regarded as two groups of three factors each; in each group we have α, β, γ, in rotation.

By contrast, in an m. i. p. expansion we have k groups of n factors each, in this illustration $\alpha_{11}\alpha_{22}\beta_{11}\beta_{22}\gamma_{11}\gamma_{22}$. The order of α's, β's, and γ's, considered separately, is the same in both cases.

Under this definition of direct expansion, we may at once

obtain from theorems I, Ia, and II, three new theorems by interchanging the multiple and the direct character of the development on both sides of the identity. It will be sufficient to re-state theorem II, which I shall do, for variety, as a multiplication theorem rather than a factor theorem:

Theorem VI. Given a determinant A whose elements are of the form $\alpha_{mi_1 \ldots i_p}$ having $p+1$ directions, and a second determinant B having $q+1$ directions and elements $\beta_{mj_1 \ldots j_q}$; the multiple indeterminate product of A and B is equal to a *direct* expansion of a determinant C having dyad elements $\alpha_{mi_1 \ldots i_p}\beta_{mj_1 \ldots j_q}$ and class $p+q+1$, provided that

1°. The paired index m is signant in C only when it is signant in one and only one of the determinants A and B.

2°. The signancy of the other indices is continued.

3°. The ordinal character of the indices in A and in B is continued.

It may be noted that the reciprocal character of the development on the two sides of the identity $AB \doteqdot C$, — *multiple on one side when direct on the other*, — shown by theorems II and VI, is a consequence of the fact that, by definition, each element of C contains both an element of A and an element of B.

Part II. Application to Polyadics

9. Definition and elementary properties of polyadics.

Let there be symbols $\mathbf{a}_{ij}$ not commutative in multiplication which may be called vectors. Let each $\mathbf{a}$ be linearly expressible in terms of n basis vectors as

$$\mathbf{a}_{ij} = \mathbf{e}_m a_{ijm} \quad \text{(summed for } m = 1, 2, \ldots n) \tag{14}$$

A polyadic $\mathbf{P}$ of degree p is defined as a sum of terms

$$\mathbf{P} = \mathbf{a}_{1j}\mathbf{a}_{2j} \ldots \mathbf{a}_{pj} \quad \text{(summed for } j = 1, 2, \ldots h) \tag{15}$$

The number of terms h will be called the *extent* of the polyadic. The first index distinguishes the first factors, second factors, etc., of the respective polyad terms.

It is *not* assumed, for the present at least, that the basis vectors $\mathbf{e}_1 \ldots \mathbf{e}_n$ possess any unit or orthogonal character. Polyadics

defined by (15) are thus consistent with the polyadics of Gibbs but not precisely identical with them. In so far as we conceive of the vectors **a** as in a *space* of n dimensions this space is affine rather than Euclidean. The developments which follow have algebraic rather than geometrical interest, geometric language, when used, being for convenience.

If every **a** in (15) be developed as in (14) and terms collected the polyadic **P** takes the form

$$\mathbf{P}=\mathbf{e}_{i_1}\mathbf{e}_{i_2}\ldots\mathbf{e}_{i_p}\,a_{i_1i_2\ldots i_p} \quad \text{(summed on all indices from 1 to } n\text{)} \quad (16)$$

where the right side contains n^p terms. The n^p scalars $a_{i_1i_2\ldots i_p}$ constitute a p-way matrix of order n. When expanded as in (16) the polyadic may be said to be *developed on its scalar components*. Arithmetically the n^p scalars $a_{i_1i_2\ldots i_p}$ of (16) are connected with the hnp scalars a_{ijm} of (14) and (15) by the n^p equations

$$a_{i_1i_2\ldots i_p}=a_{1\,ji_1}a_{2ji_2}\ldots a_{pji_p} \quad \text{(summed for } j=1,\,2,\,\ldots\,h\text{)} \quad (17)$$

A polyadic determined by a sum of polyads on the one hand, or, on the other by a scalar matrix, represents in a sense two contrasted view-points. The properties to be studied in this paper pertain to both these concepts. For fundamental to the applications which it is desired to make are certain expansions intermediate in character between the definition (15) and the scalar development (16). These expansions are obtained by expanding some of the **a**'s by aid of (14), and not expanding the rest. For example, let all $\mathbf{a}_{ij}$ be expanded except the $\mathbf{a}_{pj}$, that is to say, expand every factor of every term of the polyadic with exception of the last factor in each term. Collect the terms which explicitly contain the same basis vectors, giving the polyadic in the form

$$\mathbf{P}=\mathbf{e}_{i_1}\mathbf{e}_{i_2}\ldots\mathbf{e}_{i_{p-1}}\,a_{i_1i_2\ldots i_{p-1}} \quad \text{(summed on all indices from 1 to } n\text{)} \quad (18)$$

where the n^{p-1} *vectors* which result from collecting terms satisfy

$$a_{i_1i_2\ldots i_{p-1}}=a_{1ji_1}\,a_{2ji_2}\ldots a_{(p-1)ji_{p-1}}\mathbf{a}_{pj} \quad \text{(summed for } j=1,\,2,\,\ldots\,h\text{)} \quad (19)$$

The expansion (18) may be called the *vector expansion on the last factor*. The vectors **a** constitute a $(p-1)$-way matrix with vector elements. Similarly we may have vector expansions on any other factor.

Again, let every factor in every term of (15) be expanded by (14) with exception of the last *two* factors in each term. Then $\mathbf{P}=\mathbf{e}_{i_1}\mathbf{e}_{i_2}\ldots\ \mathbf{e}_{i_{p-2}}\ \phi_{i_1 i_2\ldots i_{p-2}}$ (summed on all indices from 1 to n) (20) where the n^{p-2} *dyadics* which result from collecting terms satisfy

$$\phi_{i_1 i_2\ldots i_{p-2}}=a_{1j i_1}a_{2 j_2}\ldots a_{(p-2)\,j i_{p-2}}\,\mathbf{a}_{(p-1)j}\mathbf{a}_{pj} \qquad (21)$$

where we sum from 1 to h on the index j in each case. We may call (20) the *dyadic expansion on the last two factors.*

In a similar way we may have a *k-adic expansion* on any k assigned factors. For the sake of collecting terms, these factors will usually be assigned adjacent to one another. If not, they can be made so by taking a *conjugate* of the polyadic in the sense of Art. 8. In general a polyadic may be determined by an n^{p-k}-way matrix whose elements are k-adics associated with k assigned factors of the polyadic. Here k may be any integer from 1 to $p-1$. In this paper the k-adic expansion will always be made on the *last* k factors. Then the polyadic takes the form

$$\mathbf{P}=\mathbf{e}_{i_1}\mathbf{e}_{i_2}\ldots\ \mathbf{e}_{i_{p-k}}\,\mathbf{M}_{i_1 i_2\ldots i_{p-k}} \qquad (21')$$

summed on all indices from 1 to n. The n^{p-k} quantities $\mathbf{M}$ are k-adics. Each $\mathbf{M}$ is related to (14) and (15) by an equation of form similar to the special cases (19) and (21). More important however is the relation of each $\mathbf{M}$ to the n^p scalars $a_{i_1 i_2\ldots i_p}$. If each $\mathbf{M}$ be developed on its scalar elements we have

$$\mathbf{M}_{i_1 i_2\ldots i_{p-k}}=\mathbf{e}_{i_{p-k+1}}\,\mathbf{e}_{i_{p-k+2}}\ldots\ \mathbf{e}_{i_p}a_{i_1 i_2\ldots i_p} \qquad (21'')$$

where the summation is of course on the last k indices from 1 to n, the first p-k indices being fixed to denote a particular $\mathbf{M}$. Thus any one k-adic occurring in a k-adic expansion of a polyadic $\mathbf{P}$ corresponds to a k-way block of scalar elements cut from the p-way scalar matrix of $\mathbf{P}$.

The largest value of k is $p-1$. In this case the polyadic takes the form

$$\mathbf{P}=\mathbf{e}_1\mathbf{M}_1+\mathbf{e}_2\mathbf{M}_2+\ \ldots\ +\mathbf{e}_n\mathbf{M}_n \qquad (22)$$

assuming the $\mathbf{M}$'s to be $(p-1)$-adics associated with the p-1 latter factors of the polyadic by the equations

$$\mathbf{M}_m=a_{1jm}\mathbf{a}_{2j}\,\mathbf{a}_{3j}\ldots\mathbf{a}_{pj}\ \text{(summed for } j=1,\,2\ldots h). \qquad (23)$$

The $\mathbf{M}_m$ of (23) are also determined by the layers of the first direction of the p-way scalar matrix.

10. **Determinants obtained by folding k-adic expansions.**

In a recent paper by Mr. L. H. Rice and the author (this Journal, Vol. IV, No 3, page 179) it was shown that if two or more polyadics are first multipled by m. i. p. and each group of corresponding factors is then replaced by its space-complement (a process called *folding*), the result is a determinant with the following construction:

1°. If polyadics of degree p be folded, the result is a determinant of class $p+1$.

2°. The first $n-q$ layers are alike, and have elements $\mathbf{e}_{i_1}\mathbf{e}_{i_2}\ldots\mathbf{e}_{i_p}$ that is p-ad elements whose factors are basis vectors. Here q is the number of polyadics folded, not greater than n.

3°. The remaining q layers consist of the n^p scalar elements of the respective polyadics, in the order of multiplication.

The applications which follow are based on the following fact: the proof given in the paper referred to is independent of commutative operations in multiplication; it therefore applies without change to the k-adic expansions above defined, — so long as these expansions are on the *latter* factors of the polyadics concerned, and provided also that the first direction of the resulting determinant (corresponding to the order of multiplication) is maintained as an ordinal direction. The k-adic coefficients will take the place of scalar coefficients. Folding will take place on the first $p-k$ factors only. The last k factors, namely those corresponding to the k-adic coefficients, will be multiplied by m. i. p. but no space-complements will be taken on these k factors. This result may be stated as

Theorem VII. Let a polyadic $\mathbf{P}$ of degree p be expanded in the form

$$\mathbf{P}=\mathbf{e}_{i_1}\mathbf{e}_{i_2}\ldots\mathbf{e}_{i_{p-k}}\mathbf{M}_{i_1 i_2\ldots i_{p-k}} \text{ (summed on all indices from 1 to } n) \tag{24}$$

where the $\mathbf{M}$'s are polyadics of degree k. Let there be $q-1$ other polyadics of degree p similarly expanded ($q \leqq n$). Let the multiple indeterminate product of these q polyadics of degree p be formed. After the multiplication let each group of q first factors be replaced by its space-complement, and similarly each group of q second

factors, and so on up to each group of q $(p-k)^{\text{th}}$ factors. The remaining groups of q factors each are left unreplaced. In short the q polyadics are *folded* on their first $p-k$ factors. The result is equal to a determinant of $p-k+1$ directions constructed as follows:

1°. The first $n-q$ layers are alike and have $(p-k)$-ad elements $\mathbf{e}_{i_1}\mathbf{e}_{i_2}\cdots\mathbf{e}_{i_{p-k}}$.

2°. The remaining q layers consist, layer by layer, of the respective sets of polyadics of degree k arising from the k-adic expansion of the given q polyadics; these layers occur in the order of multiplication of the given polyadics.

3°. The development is by m. i. p. (See remark below.)

4°. The order of multiplication of the given polyadics governs the development, *i.e.*, the first direction is ordinal for all factors.

To prove this theorem would be merely to repeat the paper referred to, replacing scalar coefficients by the k-adics $\mathbf{M}$.

Remark on the m. i. p. development. The term m. i. p is here free from ambiguity if we think of the given polyadics; namely, all factors which are, or are derived from, first factors of the given polyadics precede all which pertain to second factors, and so on. If we think of the determinant alone, we should remember that all k-adic elements come last. Thus, as it were, the basis vectors of the first $n-q$ layers follow an m. i. p. development in themselves, while the k-adic elements follow a separate, and subsequent m. i. p. development.

The simplest case which fully illustrates these facts is given by a pair of tetradics in 4-space (see (30) and (31) below). The ordered expansion of (31) is sufficiently illustrated by the leading term, which, if the expansion were *direct* would be

$$\mathbf{e}_1\mathbf{e}_1\mathbf{e}_2\mathbf{e}_2\phi_{23}\theta_{44}$$

but since the expansion must be by m. i. p becomes

$$\mathbf{e}_1\mathbf{e}_2\mathbf{e}_1\mathbf{e}_2\,[\phi_{23}\theta_{44}]$$

where, as a temporary notation, square brackets indicate the multiple indeterminate product of ϕ_{23} into θ_{44}. This example is taken up in more detail below.

We may next, however, consider two examples of theorem VII so simple that all elements of the required determinant are vectors,

and where, consequently, the question of m. i. p. expansion does not arise. Consider two ordinary dyadics in 3-space.

Folding them on both factors gives the double cross product of Gibbs. Folding on the first factor only is the same as taking their double product with cross on one factor and a stroke, for indeterminate product, on the other factor. If ϕ and θ are the dyadics expanded as

$$\phi = \mathbf{e}_1\alpha_1 + \mathbf{e}_2\alpha_2 + \mathbf{e}_3\alpha_3, \qquad \theta = \mathbf{e}_1\beta_1 + \mathbf{e}_2\beta_2 + \mathbf{e}_3\beta_3 \tag{25}$$

the result of folding on the first factor only may be indicated by $\phi_1^{\times}\theta$ and, by theorem VII, is equal to the square determinant

$$\begin{vmatrix} \mathbf{e}_1, & \mathbf{e}_2, & \mathbf{e}_3 \\ \alpha_1, & \alpha_2, & \alpha_3 \\ \beta_1, & \beta_2, & \beta_3 \end{vmatrix} \tag{26}$$

where the vertical direction is ordinal. Here $n=3$, $p=2$, $q=2$, and $k=1$.

Taking as a second illustration the case $n=3$, $p=3$, $q=3$, and $k=1$, we shall have three triadics $\mathbf{e}_i\mathbf{e}_j\alpha_{ij}$, $\mathbf{e}_i\mathbf{e}_j\beta_{ij}$, and $\mathbf{e}_i\mathbf{e}_j\gamma_{ij}$. Folding on the first two factors gives the cubic determinant

$$\begin{vmatrix} \alpha_{11}, & \alpha_{12}, & \alpha_{13} \\ \alpha_{21}, & \alpha_{22}, & \alpha_{23} \\ \alpha_{31}, & \alpha_{32}, & \alpha_{33} \end{vmatrix} \begin{vmatrix} \beta_{11}, & \beta_{12}, & \beta_{13} \\ \beta_{21}, & \beta_{22}, & \beta_{23} \\ \beta_{31}, & \beta_{32}, & \beta_{33} \end{vmatrix} \begin{vmatrix} \gamma_{11}, & \gamma_{12}, & \gamma_{13} \\ \gamma_{21}, & \gamma_{22}, & \gamma_{23} \\ \gamma_{31}, & \gamma_{32}, & \gamma_{33} \end{vmatrix} \tag{27}$$

On the other hand, if we fold only the first two of these triadics (so $q=2$ *instead of* $q=3$) we have

$$\begin{vmatrix} \mathbf{e}_1\mathbf{e}_1, & \mathbf{e}_1\mathbf{e}_2, & \mathbf{e}_1\mathbf{e}_3 \\ \mathbf{e}_2\mathbf{e}_1, & \mathbf{e}_2\mathbf{e}_2, & \mathbf{e}_2\mathbf{e}_3 \\ \mathbf{e}_3\mathbf{e}_1, & \mathbf{e}_3\mathbf{e}_2, & \mathbf{e}_3\mathbf{e}_3 \end{vmatrix} \begin{vmatrix} \alpha_{11}, & \alpha_{12}, & \alpha_{13} \\ \alpha_{21}, & \alpha_{22}, & \alpha_{23} \\ \alpha_{31}, & \alpha_{32}, & \alpha_{33} \end{vmatrix} \begin{vmatrix} \beta_{11}, & \beta_{12}, & \beta_{13} \\ \beta_{21}, & \beta_{22}, & \beta_{23} \\ \beta_{31}, & \beta_{32}, & \beta_{33} \end{vmatrix} \tag{28}$$

In both (27) and (28) the ordinal direction is non-signant, hence if the triadics are made equal to each other neither of these results vanishes identically. The leading term of (28) is $\mathbf{e}_1\mathbf{e}_1\alpha_{22}\beta_{33}$.

To summarize our results thus far we may say that *by folding q polyadics of degree p in n-space on their k-adic expansions, the result is a polyadic of degree* $(p-k)(n-q)+kq$ *which may be written as a determinant.* Constructing this determinant implies developing the folded factors in terms of basis vectors. *But it is of the utmost importance to notice that the folding process itself makes no such implication.* For it consists of two steps, multiplying by

m. i. p. and taking space-complements, both of which are known to be invariant operations in the sense that the result is independent of the manner in which the polyadics are expanded.

A suitable notation is desirable. Temporarily I shall denote the result of folding several polyadics $\mathbf{P}$, $\mathbf{Q}$, etc. on their first $p-k$ factors by $\mathrm{F}_k((\mathbf{P}, \mathbf{Q}, \ldots))$ or, merely by $((\mathbf{P}, \mathbf{Q}, \ldots))_k$, leaving the values of n and of p to be made clear from the context. Thus (26) is $\mathrm{F}_1((\phi, \theta))$ or merely $((\phi, \theta))_1$. The paper mentioned at the head of Art. 10 deals with the operation F_0, since $k=0$ when all factors are folded. There is no logical reason to limit folding to the *first* $p-k$ factors. Any other case may be dealt with either by permuting the factors of the given polyadics into the desired position, which is equivalent to permuting the indices of their scalar elements; or by explicitly designating the folded factors.

As a further example, we may quite well have $q=1$. Thus if $\phi = \mathbf{e}_1\alpha_1 + \mathbf{e}_2\alpha_2 + \mathbf{e}_3\alpha_3$ we have

$$\mathrm{F}_1((\phi)) = \begin{vmatrix} \mathbf{e}_1, & \mathbf{e}_2, & \mathbf{e}_3 \\ \mathbf{e}_1, & \mathbf{e}_2, & \mathbf{e}_3 \\ \alpha_1, & \alpha_2, & \alpha_3 \end{vmatrix} \tag{29}$$

11. The scalar components of the polyadics $\mathbf{F}_k$.

I shall now show that the *scalar components* of the polyadics $\mathbf{F}_k$ obtained in the last article are themselves determinants, and shall exhibit their form. If $\mathbf{F}_k$ be written in the determinant form described under theorem VII let it be called D_k. Now D_k has $n-q$ layers constructed from *basis* vectors and q layers whose elements are k-adics. Let D_k be expressed as a sum of products, by means of a Laplacean expansion, in the following manner: each product shall be the *direct* product of two determinants, the first an $(n-q)$-rowed determinant made out of basis vectors, the second a q-rowed determinant with k-adic elements. The simplest illustration is afforded by a pair of tetradics in four-space folded on the first two factors, that is $n=4$, $p=4$, $q=2$, and $k=2$. Let tetradics $\mathbf{P}$ and $\mathbf{P}'$ be expanded on the last two factors thus,

$$\mathbf{P} = \mathbf{e}_i\mathbf{e}_j\phi_{ij} \text{ and } \mathbf{P}' = \mathbf{e}_i\mathbf{e}_j\theta_{ij} \text{ (summed for } i, j, = 1, 2, 3, 4) \tag{30}$$

where the ϕ_{ij} and θ_{ij} are two sets of dyadic coefficients. By

theorem VII the polyadic $\mathbf{F}_2((\mathbf{P}, \mathbf{P}'))$ is equal to the m. i. p. expansion of the cubic determinant.

$$\begin{vmatrix} \mathbf{e}_1\mathbf{e}_1, & \mathbf{e}_1\mathbf{e}_2, & \mathbf{e}_1\mathbf{e}_3, & \mathbf{e}_1\mathbf{e}_4 \\ \mathbf{e}_2\mathbf{e}_1, & \mathbf{e}_2\mathbf{e}_2, & \mathbf{e}_2\mathbf{e}_3, & \mathbf{e}_2\mathbf{e}_4 \\ \mathbf{e}_3\mathbf{e}_1, & \mathbf{e}_3\mathbf{e}_2, & \mathbf{e}_3\mathbf{e}_3, & \mathbf{e}_3\mathbf{e}_4 \\ \mathbf{e}_4\mathbf{e}_1, & \mathbf{e}_4\mathbf{e}_2, & \mathbf{e}_4\mathbf{e}_3, & \mathbf{e}_4\mathbf{e}_4 \end{vmatrix} \begin{vmatrix} \mathbf{e}_1\mathbf{e}_1, & \mathbf{e}_1\mathbf{e}_2, & \mathbf{e}_1\mathbf{e}_3, & \mathbf{e}_1\mathbf{e}_4 \\ \mathbf{e}_2\mathbf{e}_1, & \mathbf{e}_2\mathbf{e}_2, & \mathbf{e}_2\mathbf{e}_3, & \mathbf{e}_2\mathbf{e}_4 \\ \mathbf{e}_3\mathbf{e}_1, & \mathbf{e}_3\mathbf{e}_2, & \mathbf{e}_3\mathbf{e}_3, & \mathbf{e}_3\mathbf{e}_4 \\ \mathbf{e}_4\mathbf{e}_1, & \mathbf{e}_4\mathbf{e}_2, & \mathbf{e}_4\mathbf{e}_3, & \mathbf{e}_4\mathbf{e}_4 \end{vmatrix} \begin{vmatrix} \phi_{11}, & \phi_{12}, & \phi_{13}, & \phi_{14} \\ \phi_{21}, & \phi_{22}, & \phi_{23}, & \phi_{24} \\ \phi_{31}, & \phi_{32}, & \phi_{33}, & \phi_{34} \\ \phi_{41}, & \phi_{42}, & \phi_{43}, & \phi_{44} \end{vmatrix} \begin{vmatrix} \theta_{11}, & \theta_{12}, & \theta_{13}, & \theta_{14} \\ \theta_{21}, & \theta_{22}, & \theta_{23}, & \theta_{24} \\ \theta_{31}, & \theta_{32}, & \theta_{33}, & \theta_{34} \\ \theta_{41}, & \theta_{42}, & \theta_{43}, & \theta_{44} \end{vmatrix} \tag{31}$$

where the square layers, as written, are layers of the first, or non-signant direction, ordinal for all factors. In the Laplacean expansion under consideration the leading term is the product (*direct* because the first two layers pertain to the first two factors of the tetradics) of the two-row cubic determinant with dyad elements

$$\begin{vmatrix} \mathbf{e}_1\mathbf{e}_1, & \mathbf{e}_1\mathbf{e}_2 \\ \mathbf{e}_2\mathbf{e}_1, & \mathbf{e}_2\mathbf{e}_2 \end{vmatrix} \begin{vmatrix} \mathbf{e}_1\mathbf{e}_1, & \mathbf{e}_1\mathbf{e}_2 \\ \mathbf{e}_2\mathbf{e}_1, & \mathbf{e}_2\mathbf{e}_2 \end{vmatrix} \tag{32}$$

nto the two-row cubic determinant with *dyadic* elements

$$\begin{vmatrix} \phi_{33}, & \phi_{34} \\ \phi_{43}, & \phi_{44} \end{vmatrix} \begin{vmatrix} \theta_{33}, & \theta_{34} \\ \theta_{43}, & \theta_{44} \end{vmatrix} \tag{33}$$

and the other terms are of like character with proper sign. Since configurations like (32) contain basis vectors only, the scalar elements of $\mathbf{F}_2$ will be scalar elements of the m. i. p. expansion of (33) and the like. Now we may write

$$\phi_{ij} = \mathbf{e}_r\mathbf{e}_s a_{ijrs} \text{ and } \theta_{ij} = \mathbf{e}_x\mathbf{e}_y b_{ijxy} \tag{34}$$

whence the scalar coefficient of $\mathbf{e}_r\mathbf{e}_s\mathbf{e}_x\mathbf{e}_y$ in the m. i. p. expansion of any determinant of the form of (33) is

$$\begin{vmatrix} a_{ACrs}, & a_{ADrs} \\ a_{BCrs}, & a_{BDrs} \end{vmatrix} \begin{vmatrix} b_{ACxy}, & b_{ADxy} \\ b_{BCxy}, & b_{BDxy} \end{vmatrix} \tag{35}$$

where in the special case of (33) we have $A=C=3$ and $B=D=4$. The method being perfectly general, the scalar components of any $\mathbf{F}_k$ must be determinants obtained in a similar manner. The result may be specifically stated as follows: *the scalar components of the polyadic* $\mathbf{F}_k((\mathbf{P}, \mathbf{P}', \mathbf{P}'', \ldots))$ *are scalar* $(p-k+1)$*-way determinants of order* q*; the elements of its* q *layers are scalar components of the respective* q *polyadics* $\mathbf{P}, \mathbf{P}', \mathbf{P}'', \ldots$*; the last* k *indices of these elements are constant over any one layer, while the variation of the first* $p-k$ *indices gives the* $p-k$ *directions within the layer.*

This rule gives the form of each scalar component of $\mathbf{F}_k$. We have now to determine the selection of basis vectors in the term of $\mathbf{F}_k$ to which it belongs, and the sign to be prefixed. Returning to (32), it follows from Part I that all such configurations of basis vectors break up into factors. For if we append to every $\mathbf{e}$ a dummy index to denote identical layers, such a determinant comes under theorem II. By successive applications of this theorem when necessary, we shall have the direct product of a set of factors each of which is an $(n-q)$-rowed minor of the fundamental determinant of basis vectors

$$\begin{vmatrix} \mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n \\ \mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n \\ \cdots\cdots\cdots \\ \mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n \end{vmatrix} \tag{36}$$

and there will be $p-k$ such factors. Thus the m. i. p. expansion of (32) breaks up into $(\mathbf{e}_1\mathbf{e}_2-\mathbf{e}_2\mathbf{e}_1)$ $(\mathbf{e}_1\mathbf{e}_2-\mathbf{e}_2\mathbf{e}_1)$ and the m. i. p. expansion of

$$\begin{vmatrix} \mathbf{e}_a\mathbf{e}_c, \mathbf{e}_a\mathbf{e}_d \\ \mathbf{e}_b\mathbf{e}_c, \mathbf{e}_b\mathbf{e}_d \end{vmatrix} \begin{vmatrix} \mathbf{e}_a\mathbf{e}_c, \mathbf{e}_a\mathbf{e}_d \\ \mathbf{e}_b\mathbf{e}_c, \mathbf{e}_b\mathbf{e}_d \end{vmatrix} \tag{37}$$

breaks up into $(\mathbf{e}_a\mathbf{e}_b-\mathbf{e}_b\mathbf{e}_a)$ $(\mathbf{e}_c\mathbf{e}_d-\mathbf{e}_d\mathbf{e}_c)$, the non-signant direction being ordinal for both factors in agreement with theorem VII. Now the minors of (36) may with some reason be regarded as new fundamental units. Whether we view them in that light or not, it will be well for brevity to denote them by the symbol $\mathbf{E}$ with subscripts corresponding with their leading term, (32) becoming $\mathbf{E}_{12}\mathbf{E}_{12}$ and (37) becoming $\mathbf{E}_{ab}\mathbf{E}_{cd}$. In general then, the determinants with $(p-k)$-ad elements which occur in the Laplacean expansion of F_k break up into the direct product of $p-k$ quantities $\mathbf{E}$, each carrying $n-p$ subscripts. These subscripts are the index numbers of the files in which the determinants lie. Thus (37) lies in rows a and b and in columns c and d in the first square layer of (31); the ordinal index, being a dummy, need not be written.

We can now state the relation connecting any chosen scalar component of $\mathbf{F}_k$ with the polyads to which it belongs: *any chosen scalar component of* $\mathbf{F}_k$ occurs not in one term merely, *but is the coefficient of a polyadic; this polyadic is the direct product of* $p-k$

quantities **E**, *each carrying* $n-q$ *indices, into* kq *basis vectors* **e**; *the indices of the* **E**'*s are complementary to the first* $p-k$ *indices of the scalar and the indices of the* **e**'*s are the same as the other indices of the scalar.*

Remark on the words "complementary" and "same" in the above rule: to say that the indices carried by a particular **E** are complementary to the corresponding index of an 'a' in the scalar component means that the index of 'a' shall take on all values *except* those (fixed) values possessed by the indices of **E**: (example: the first index in (33) takes on values 3 and 4 while the first factor in (32) $\mathbf{E}_{12}$ carries the complementary indices 1 and 2); on the other hand in the expansion of (33) the indices r, x, s, and y in $\mathbf{e}_r\mathbf{e}_x\mathbf{e}_s\mathbf{e}_y$ are the *same* as the last k (fixed) indices of the q layers of (35), *in the order defined by m. i. p. expansion.*

This relation of indices will naturally be of importance if we wish later to study the invariant properties of $\mathbf{F}_k$. No apology will be needed, therefore, if a special notation be introduced. I shall accordingly adopt small italic indices of the first part of the alphabet for the **E**'s (compare (37), and small capitals for the corresponding indices of the a's. Italics of the latter part of the alphabet will be reserved for the basis vectors **e** and the corresponding indices of the a's. The *italic indices* will have *fixed* values in any one scalar component and its polyadic coefficient, while the small caps will take on q different values as we pass along a file of the determinant (compare (35)). Thus any one scalar component of $\mathbf{F}_k$, will multiply the polyadic

$$\mathbf{E}^{(1)}_{a_1\ldots a_{n-q}}\mathbf{E}^{(2)}_{b_1\ldots b_{n-q}}\cdots\mathbf{E}^{(p-k)}_{g_1\ldots g_{n-q}}\mathbf{e}^{(1)}_{r_1}\mathbf{e}^{(1)}_{r_2}\cdots\mathbf{e}^{(1)}_{r_q}\mathbf{e}^{(2)}_{s_1}\cdots\mathbf{e}^{(2)}_{s_q}\cdots\mathbf{e}^{(k)}_{z_1}\mathbf{e}^{(k)}_{z_2}\cdots\mathbf{e}^{(k)}_{z_q} \tag{38}$$

where the scalar component itself will be the determinant whose elements are

$$a^{(i)}_{\mathrm{AB}}\cdots\alpha_{r_i s_i}\cdots z_i \tag{39}$$

Remembering that this scalar component is a determinant of order q we see that its elements of any layer are obtained by varying each of the indices A, B, ... G over the q values *not* found in the sets a, b, . . . g, respectively; while the q ordinal layers are obtained by varying i from 1 to q. Thus the scalar

determinant is of class $p-k+1$ as already stated. Instead of $a^{(i)}$ we may, if preferred, use $a, b, \ldots$ to q elements.

Finally, to determine the sign to be prefixed, it is only necessary to notice that the $n-q$ integers $a_1, a_2, \ldots a_{n-q}$, together with the q values taken on by A, determine a *range*, to which Rice's rule applies, and similarly for $b, c, \ldots g$. These $p-k$ ranges determine the sign of the result, for the determinant D_k has $p-k+1$ directions, but the ordinal direction may be neglected in applying Rice's rule.

Furthermore it is only necessary to consider the leading term of (39) and the leading term of each $\mathbf{E}$, that is, we may assume the indices attached to any $\mathbf{E}$ to stand in ascending order of numerical value, and consider only the main diagonal of the scalar determinant. (The indices $r, s, \ldots z$ do not affect the sign in any way.) Rice's rule may then be applied.

As an example of the direct application of (38) and (39), let it be required to write down the scalar coefficient of

$$\mathbf{E}_{12}\mathbf{E}_{23}\mathbf{e}_4\mathbf{e}_3\mathbf{e}_2 \tag{40}$$

in the Laplacean development of (31). Aside from sign, the above rules give

$$\begin{vmatrix} a_{2143}, & a_{2443} \\ a_{4143}, & a_{4443} \end{vmatrix} \begin{vmatrix} b_{2132}, & b_{2432} \\ b_{4132}, & b_{4432} \end{vmatrix} \tag{41}$$

where the first index is complementary to $\mathbf{E}_{12}$, the second is complementary to $\mathbf{E}_{23}$, the third indices in the two layers are respectively 4 and 3, and the fourth indices are 3 and 2 in agreement with the configuration $\mathbf{e}_4\mathbf{e}_3\mathbf{e}_2\mathbf{e}_2$. To find the sign, we note the two signant ranges 13 .4 and 2314 for the leading term, hence the sign is minus.

As another example, let the coefficient of $\mathbf{E}_1\mathbf{E}_3\mathbf{e}_2\mathbf{e}_3$ be required from (28). We note first that, formally, $\mathbf{E}_1 = \mathbf{e}_1$ and $\mathbf{E}_3 = \mathbf{e}_3$. The rules give

$$\begin{vmatrix} a_{212}, & a_{222} \\ a_{312}, & a_{322} \end{vmatrix} \begin{vmatrix} b_{213}, & b_{223} \\ b_{313}, & b_{323} \end{vmatrix} \tag{42}$$

with signant ranges 123 and 312 so the sign is plus.

In cases where $q=n$ there are no $\mathbf{E}$'s. The first $p-k$ indices in any scalar component then range from 1 to n Thus the coeffi-

cient of $\mathbf{e}_r\mathbf{e}_s\mathbf{e}_t$ in (27) is the cubic determinant whose leading term is $a_{1r}b_{2s}c_{3t}$.

The expressions (38) and (39) determine the "general term" of any $\mathbf{F}_k$. The relation between the three sets of indices, while perhaps formidable in appearance, is in reality quite simple, as the examples show, provided the distinctions of fixed and complementary indices are held in mind. Speaking roughly, the bearing of these results may be suggested by the fact that, in applications to algebraic forms, these three sets of indices possess, respectively, contravariant, covariant, and invariant character. Most, if not all, concomitants are implicitly contained in (38) and (39).

12. **Folding a polyadic into itself.**

The polyadics $\mathbf{F}_k$ of Art. 11 are invariants of a system of polyadics. As an important special class of cases, where certain marked simplifications occur, these given polyadics may be equal to each other. A sharp distinction now arises according to whether $p-k+1$, the class of D_k, *is even or odd*. If this integer is odd the ordinal direction in D_k is non-signant. It is also non-signant in the scalar determinants which, as above shown, are the scalar components of $\mathbf{F}_k$. Therefore equality of the given q polyadics, which entails identity of layers in some of the scalar components, does not in general cause any component to vanish. But if $p-k+1$ is even, any scalar component of $\mathbf{F}_k$ will vanish which has two ordinal layers alike. Important simplifications will thereby result. I shall now consider various cases according to the value of k.

Case $k=0$. Since q layers of D_k are composed of scalar elements, namely the scalar elements of the q given polyadics, $\mathbf{F}_k$ vanishes identically when p is odd, and any two of the given polyadics are equal. Stating the fact with reference to a single polyadic we have

Theorem VIII. If a polyadic of odd degree be folded into itself on all factors the result is zero.

If we agree to write $\mathbf{F}_k((\mathbf{P}^q))$ for the result of folding q equal polyadics on their first $p-k$ factors, theorem VIII might be stated in the following alternative manner: $\mathbf{F}_0((\mathbf{P}^q))=0$ when p is odd and $q>1$.

When $k=0$ and p is even, we have generalizations of the cases most often treated by writers on vector algebra, where, usually, $p=2$, $n=3$. Thus for dyadics in three-space, $\mathbf{F}_0((\mathbf{P}^2))$ is the double-cross " square," and $\mathrm{F}_0((\mathbf{P}^3))$ in Gibbs' notation would be $\mathbf{P}:\mathbf{P}\overset{\times}{\times}\mathbf{P}$. More generally, if p is even $\mathrm{F}_0((\mathbf{P}^n))$ may by therorem VII be written as a determinant with n layers in the non-signant direction alike, and these layers have for their elements the scalar components of the given polyadic $\mathbf{P}$, whence

Theorem IX. When p is even, $\mathrm{F}_0((\mathbf{P}^n))$ is equal to $n!$ times the full sign determinant of the scalar components of $\mathbf{P}$.

If we now apply theorem I we may show

Theorem X. If a polyadic $\mathbf{P}$ is the direct product of two polyadics $\mathbf{A}$ and $\mathbf{B}$, then $\mathrm{F}_0((\mathbf{P}^n))=\mathrm{F}_0((\mathbf{A}^n))\mathrm{F}_0((\mathbf{B}^n))$.

For by theorems I and IX (in case p is even) $\mathrm{F}_0((\mathbf{P}^n))$ is $(n!)^2$ times the product of the full sign determinants of the components of $\mathbf{A}$ and $\mathbf{B}$, whence, applying theorem IX to $\mathbf{A}$ and $\mathbf{B}$, (in case they are both of even degree), $\mathrm{F}_0((\mathbf{P}^n))$ is the product of $\mathrm{F}_0((\mathbf{A}^n))$ and $\mathrm{F}_0((\mathbf{B}^n))$.

In case p is odd, the theorem is true but trivial, for either $\mathbf{A}$ or $\mathbf{B}$ must be of odd degree and both sides of the equation are zero by theorem VIII.

In case p is even while $\mathbf{A}$ and $\mathbf{B}$ are both of odd degree, the scalar components of $\mathbf{P}$ are of the form contemplated in theorem III. The full sign determinant of these components therefore vanishes. This result may be stated as

Theorem XI. If a polyadic $\mathbf{P}$ is the direct product of two polyadics $\mathbf{A}$ and $\mathbf{B}$ of odd degree, the full sign determinant of the components of $\mathbf{P}$ vanishes.

A familiar special case is the vanishing of the determinant of a dyadic when the dyadic reduces to a single dyad. The vanishing of the determinant is not, however, a *sufficient* condition for factoring, as witnessed by this same simple case, — the determinant of a dyadic in three-space vanishes when the dyadic reduces to the sum of *two* dyads. A complete discussion of factoring lies beyond the limits of this paper.

Case $k=1$. In the case of $\mathbf{F}_1((\mathbf{P}^q))$ the situation is again quite different according as p is odd or even. I shall next prove

Theorem XII. If $\mathbf{P}$ is a polyadic of odd degree, $\mathbf{F}_1((\mathbf{P}^n))$ is a *symmetrical* polyadic (of degree n).

Proof. We have $p-k+1=p$, an odd integer by hypothesis. Since $q=n$ the ordinal layers of D_k are all alike. The elements of any ordinal layer are vectors, namely the n^{p-1} vector coefficients of (18). If we expand D_k on its scalar components as in Art. 11, no $\mathbf{E}$'s will appear. Any term contains n basis vectors. Since the ordinal direction is non-signant, and since a scalar determinant is unaltered by interchanging non-signant layers, the scalar coefficients of all terms containing the same basis vectors will be the same, irrespective of the order of these basis vectors. But this is precisely the definition of a symmetrical polyadic. The theorem is therefore proved. To take the simplest example, let $\mathbf{P}$ be a triadic in two dimensions,

$$\mathbf{P}=\mathbf{e}_1\mathbf{e}_1a_{11}+\mathbf{e}_1\mathbf{e}_2a_{12}+\mathbf{e}_2\mathbf{e}_1a_{21}+\mathbf{e}_2\mathbf{e}_2a_{22} \tag{43}$$

Then $\mathbf{F}_1$ may be written in the form of the cubic determinant

$$\begin{vmatrix} a_{11} & a_{12} & a_{11} & a_{12} \\ a_{21} & a_{22} & a_{21} & a_{22} \end{vmatrix} \tag{44}$$

which develops into

$$2\mathbf{e}_1\mathbf{e}_1(a_{111}a_{221}-a_{211}a_{121})+(\mathbf{e}_1\mathbf{e}_2+\mathbf{e}_2\mathbf{e}_1)(a_{111}a_{222}-a_{211}a_{122}-a_{121}a_{212}+a_{221}a_{112})+2\mathbf{e}_2\mathbf{e}_2(a_{112}a_{222}-a_{212}a_{122}). \tag{45}$$

Here the coefficients of $\mathbf{e}_1\mathbf{e}_1$ and $\mathbf{e}_2\mathbf{e}_2$ might have been written as cubic determinants with two layers alike.

In this paper the notation $\mathbf{F}_1((\mathbf{P}^n))$ has been reserved to indicate folding on all factors but the last. It is evident that a similar result to theorem XII would be obtained by folding on all but one factor no matter which. The polyadics thus obtained are all symmetrical, and may be formed from $\mathbf{F}_1((\mathbf{P}^n))$ by permuting indices. Thus from (45) for example, two other polyadics may be obtained by interchanging on every "a," the first index with the third, or the second with the third. Note that (45) is unaltered by interchanging the first index with the second, because these two indices were treated alike in obtaining (45), *i.e.*, the first two factors were folded. In general the n distinct polyadics formed from $\mathbf{P}^n$(when p is odd) by folding on all but one factor are obtained by first forming $\mathbf{F}_1((\mathbf{P}^n))$ and then interchanging the last index

with each of the others. The existence of these n distinct polyadics is in part a consequence of the fact that, for a scalar matrix of odd class, there are n distinct full-sign determinants, which will, in fact, appear as the respective coefficients of $\mathbf{e}_1\mathbf{e}_2 \ldots \mathbf{e}_n$ in the n polyadics. Thus in (45) the coefficient of $(\mathbf{e}_1\mathbf{e}_2+\mathbf{e}_2\mathbf{e}_1)$ is the determinant of the components of $\mathbf{P}$ with the third index non-signant.

When p is even, $\mathbf{F}_1((\mathbf{P}^n))$ takes a very simple form,

Theorem XIII. If $\mathbf{P}$ is a polyadic of even degree, $\mathbf{F}_1((\mathbf{P}^n))$ factors into the product of the full sign determinant of the scalar components of $\mathbf{P}$ times the fundamental determinant of basis vectors (36).

Proof. All directions in D_k are signant and no $\mathbf{E}$'s appear. Any term in the expansion contains n basis vectors. If any two of these are the same the scalar coefficient is a determinant with two signant layers alike and vanishes. The coefficient of $\mathbf{e}_1\mathbf{e}_2\mathbf{e}_3 \ldots \mathbf{e}_n$ (in natural order) is the full sign determinant of the scalar components of $\mathbf{P}$, by (38) and (39). The coefficient of the same basis vectors with order interchanged in any manner gives the same determinant except that its layers in the ordinal direction (signant) are interchanged in like manner. This establishes the theorem.

Theorem XIII may, in essence, be stated without reference to polyadics, as follows:

A full-sign determinant of even class and order n, having its elements vectors in n-space and all ordinal layers alike, factors into the product of the full-sign determinant of the scalar components of its elements times the fundamental determinant (36) of basis vectors.

Theorem XIII differs from the above statement merely in the fact that the vector elements are taken as coefficient in the vector expansion of a polyadic as in (18).

This result is a generalization (to any vector determinant of even class) of the well-known fact that in three dimensions

$$\begin{vmatrix} a_1, & a_2, & a_3 \\ a_1, & a_2, & a_3 \\ a_1, & a_2, & a_3 \end{vmatrix} = \begin{vmatrix} a_{11} & a_{12}, & a_{13} \\ a_{21}, & a_{22}, & a_{23} \\ a_{31}, & a_{32}, & a_{33} \end{vmatrix} \begin{vmatrix} \mathbf{e}_1, & \mathbf{e}_2, & \mathbf{e}_3 \\ \mathbf{e}_1, & \mathbf{e}_2, & \mathbf{e}_3 \\ \mathbf{e}_1, & \mathbf{e}_2, & \mathbf{e}_3 \end{vmatrix} \tag{46}$$

Since the full sign determinant of even class is symmetrical in the p indices of the a's, it is evident that whether the factor which is not folded is the last or some other factor makes no difference in $\mathbf{F}_1$ in this case. When p is even, this full sign determinant of the scalar components is an invariant of the polyadic $\mathbf{P}$, but not when p is odd.

Theorem XIII has a bearing upon operational methods. Just as the Hamilton-Cayley equation can easily be derived from (46), so various operational equations can be obtained from polyadics of higher degree. Theorems comparable to theorems XII and XIII are easily found for higher values of k. For example, $\mathbf{F}_k$ is always *symmetrical in the factors not folded* when the q given polyadics are equal. Applications to quantics will be taken up in the succeeding paper.

A NEW METHOD IN THE THEORY OF QUANTICS

By Frank L. Hitchcock

Part I. Definitions and Elementary Theorems

1. Peculiarities of the method.

The method to be described has three new features:

1°. A non-commutative algebra is employed in which the usual symbolic processes appear as actual processes. Invariants of quantics, as well as covariants, contravariants, and mixed concomitants, appear as special cases of invariants, etc., of polyadics.

2°. Invariants, etc., are obtained by a method called *folding*[1] which is a combination of Gibbs' (multiple) indeterminate product with the (slightly generalized) space complement of Dr. Almar Naess.[2]

3°. Invariants and other concomitants appear in the form of p-way determinants.

2. Vector complexes and their correspondence.

Let polyadics be defined as in OD, Art. 9, p. 222. It is to be noted that no conception of dot product or of orthogonality enters into the work. In this paper a subscript (usually i, j, k, m, r, s, or t) entering twice in any term indicates summation from 1 to n. Let the totality of vectors $\mathbf{e}_i a_i$ be called a *complex*. Let there be a second, new, and distinct set of basis vectors $\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_n$ and a new complex comprising the totality of vectors $\mathbf{f}_i a_i'$. To any chosen vector $\mathbf{a}$ $(=\mathbf{e}_i a_i)$ of the original complex let a vector $\mathbf{a}'$ of the new complex be made to correspond by virtue of the vector equation

$$\mathbf{a}' = \mathbf{f}_i a_i' = \mathbf{f}_i \epsilon_{ji} a_j \tag{1}$$

where the determinant ϵ of the n^2 scalars ϵ_{ji} is assumed not zero.

[1] The folding process is described in two papers, *The multiple complement of one or more polyadics*, by Mr. L. H. Rice and the author, this journal, Vol. IV, No. 3, p. 179, to be referred to as "R & H," and *A theory of ordered determinants*, this journal, this volume and number, p. 205, to be referred to as "OD."

[2] See R & H, p. 179.

The first index sums with the scalar components of the chosen vector, the second with the new basis vectors.

By comparing scalar components on the two sides of (1), that is, by giving to i successive values from 1 to n instead of summing on i, we have a set of n scalar equations for the components of $\mathbf{a}'$,

$$a_i' = \epsilon_{ji} a_j \tag{2}$$

where the summation is on the first index.

By choosing the vector $\mathbf{a}$ such that $a_j = 0$ or 1 according as j is different from k or not, $\mathbf{a}$ becomes $\mathbf{e}_k$. From (1) we may thus obtain a set of n vector equations giving the correspondence of certain vectors in the new complex to the original basis vectors,

$$\mathbf{e}_k' = \mathbf{f}_i \epsilon_{ki}. \tag{3}$$

Remark on the correspondence of vectors as above defined. All accented vectors $\mathbf{a}'$, including the $\mathbf{e}_k'$, are vectors in the second complex. Furthermore the unaccented vectors $\mathbf{f}_i$ are by definition in the second complex. Accented vectors correspond to those vectors of the first complex which are denoted by the same letters without accent. It would doubtless be more consistent, formally, also to accent the new basis vectors $\mathbf{f}_i$, and this can easily be done if desired. But as there will never, or at least not at present, be any need to consider those vectors of the first complex to which correspond the basis vectors of the second complex, it is needless to accent the $\mathbf{f}_i$.

If we wish at any time to regard the two complexes as the same complex in one-one correspondence with itself, this can be done by regarding the two sets of basis vectors as identical. Again, we might regard the corresponding vectors $\mathbf{e}_k'$ as a new set of basis vectors. In the present paper they are not so regarded. In this investigation such expressions as *the components of* $\mathbf{a}'$ *with respect to the basis vectors* $\mathbf{e}_k$ have no meaning. For $\mathbf{a}'$ is not in the complex of the $\mathbf{e}_i a_i$. The components a_i' are with respect to the $\mathbf{f}_i$ by definition.

It seems hardly necessary to remark that the ϵ_{ji} are not restricted to the case of an orthogonal transformation, nor to emphasize the fact that such properties of the space complement as hold only under orthogonality of basis vectors are not now assumed. There is certainly no reason to modify the name given by Dr. Naess to

this fundamental quantity, even though its geometric significance is lost in this paper.

3. **Correspondence of polyadics.**

Let a polyadic **A** be given in any manner as a sum of polyads, all vector factors being vectors of the first complex. Let a new polyadic **A′** be formed by replacing all vector factors of the given polyadic by their corresponding vectors in the new complex according to (1). The new polyadic **A′** will be said to *correspond* to **A.** A polyadic all of whose vector factors are in the same complex will be said to " be in " that complex. The corresponding polyadic **A′** is independent of the particular manner in which the given polyadic **A** is expressed as a sum of polyads, because vectors, though not commutative with each other in multiplication, are distributive, and are commutative with scalars. In other words, the operation of replacing a polyad or polyadic by the corresponding polyads or polyadics is distributive over the terms of a sum. This property, vital to the present study, may for brevity be called *distributive invariance*, and will be said to belong to any operation which, when performed term by term upon a polyadic, gives a result independent of the particular manner in which the polyadic is expressed as a sum of polyads.

Two polyadics are equal when, and only when, they are in the same complex, and have the same scalar components each to each. If **A** be developed on its scalar components as $\mathbf{e}_{i_1}\mathbf{e}_{i_2}\cdot\cdot\cdot\mathbf{e}_{i_p}a_{i_1i_2}\cdot\cdot i_p$ and **A′** similarly as $\mathbf{f}_{i_1}\mathbf{f}_{i_2}\cdot\cdot\cdot\mathbf{f}_{i_p}a'_{i_1i_2}\cdot\cdot i_p$ then as a consequence of (1) the new polyadic **A′** is determined by the polyadic equation

$$\mathbf{A}'=\mathbf{f}_{i_1}\mathbf{f}_{i_2}\cdot\cdot\cdot\mathbf{f}_{i_p}\epsilon_{j_1i_1}\epsilon_{j_2i_2}\cdot\cdot\cdot\epsilon_{j_pi_p}a_{j_1j_2}\cdot\cdot j_p \tag{4}$$

and consequently the components of **A′** are given by the set of n^p scalar equations

$$a'_{i_1i_2}\cdot\cdot i_p=\epsilon_{j_1i_1}\epsilon_{j_2i_2}\cdot\cdot\cdot\epsilon_{j_pi_p}a_{j_1j_2}\cdot\cdot j_p \tag{5}$$

4. **Invariants of polyadics.**

Definition I. Let F be a scalar function of the scalar components of a set of given polyadics in the first complex. Let $\overline{\mathrm{F}}$ be the same function of the scalar components of the corresponding polyadics

in the second complex. If $\overline{F} = \epsilon^n F$ the quantity F is said to be an invariant of the given set of polyadics.

As an example let $n=2$ and take the binary dyadic $\mathbf{e}_i\mathbf{e}_j a_{ij}$. The quantity $a_{12}-a_{21}$ is an invariant of the dyadic. I shall give two proofs, illustrating two typical methods of procedure.

First proof. The corresponding dyadic by (4) is $\mathbf{f}_i\mathbf{f}_j a_{si}\epsilon_{si}\epsilon_{ij}$. Then

$$\begin{aligned}\overline{F} &= a_{s1}\epsilon_{s1}\epsilon_{12} - a_{s1}\epsilon_{s2}\epsilon_{11} \\ &= a_{11}\epsilon_{11}\epsilon_{12} + a_{12}\epsilon_{11}\epsilon_{22} + a_{21}\epsilon_{21}\epsilon_{12} + a_{22}\epsilon_{21}\epsilon_{22} \\ &\quad - a_{11}\epsilon_{12}\epsilon_{11} - a_{12}\epsilon_{12}\epsilon_{21} - a_{21}\epsilon_{22}\epsilon_{11} - a_{22}\epsilon_{22}\epsilon_{21} \\ &= (a_{12}-a_{21})\ (\epsilon_{11}\epsilon_{22} - \epsilon_{12}\epsilon_{21}) \\ &= \epsilon F\end{aligned}$$

Second proof. A dyadic may be regarded as a sum of dyads of the form **ab**. Let (**ab**) denote the space complement of **ab**. The space complement has the property of distributive invariance,[3] that is $\Sigma(\mathbf{ab})$ and $\Sigma(\mathbf{a'b'})$ are respectively the same functions of corresponding components. Also for each term $(\mathbf{a'b'}) = \epsilon(\mathbf{ab})$ because in this case the space complement is merely the determinant of the components of the vectors which occur. Hence $\Sigma(\mathbf{ab})$ is an invariant of the given dyadic. Taking the space complement distributively over the terms of the scalar expansion $\mathbf{e}_i\mathbf{e}_j a_{ij}$ we have $(\mathbf{e}_1\mathbf{e}_1) = (\mathbf{e}_2\mathbf{e}_2) = 0$ while $(\mathbf{e}_1\mathbf{e}_2) = 1$ and $(\mathbf{e}_2\mathbf{e}_1) = -1$. The result is thus $a_{12}-a_{21}$.

Remark on the invariant $a_{12}-a_{21}$ of a binary dyadic. This invariant is the simplest example of the fact that a polyadic has more invariants than the quantic of the same degree. It is well known that a quantic does not possess an invariant linear in its coefficients. From a polyadic its related quantic may be obtained by making the basis vectors commutative, which is the same, except for notation, as replacing the $\mathbf{e}_i$ by a set of scalars x_i. The terms $\mathbf{e}_1\mathbf{e}_2$ and $\mathbf{e}_2\mathbf{e}_1$ then cease to be distinguishable, the coefficients a_{12} and a_{21} have no separate meaning, and the invariant $a_{12}-a_{21}$ vanishes identically. On the other hand it is evident by definition that *any invariant of a polyadic or of several polyadics which does*

[3]This property of the space complement is a consequence of its definition as a determinant, independent of whether the basis vectors are regarded as forming a unit-orthogonal system. The complete proof is given by Naess (see R & H, p. 179).

not vanish identically when indices are commutative is, under this hypothesis, an invariant of the related quantic, or quantics.

It is easy to see, however, that $a_{12}-a_{21}$ is an invariant of the bilinear form $x_iy_ja_{ij}$, the proof being identical with the first proof above. Any polyadic may be made into its related multilinear form by allowing all k^{th} vector factors, for all values of k, to be commutative among themselves but not with other factors, since this is the same, except for notation, as replacing, for all values of k, the basis vectors of the k^{th} factors by a set of scalars $x_i^{(k)}$. It is then immaterial whether we distinguish the different factors by an upper index, by a change of lettering, or by an order of writing. Multilinear forms will not be further studied in this paper.

5. Invariants obtained by folding.

Polyadics are folded by first taking their multiple indeterminate product (if there be more than one polyadic) and then replacing one or more groups of corresponding factors by their space complement. In the process F_0 all such groups of factors are replaced.

Theorem I. Given n polyadics $\mathbf{A}, \mathbf{B}, \ldots \mathbf{G}$ all of degree p, the quantity $F_0((\mathbf{A}, \mathbf{B}, \ldots \mathbf{G}))$ is an invariant of these polyadics.

Proof. Let each of the given polyadics be written in any manner as a sum of polyads, $\mathbf{A}=\Sigma\mathbf{a}_1\mathbf{a}_2 \ldots \mathbf{a}_p$, $\mathbf{B}=\Sigma\mathbf{b}_1\mathbf{b}_2 \ldots \mathbf{b}_p$, etc., up to $\mathbf{G}=\Sigma\mathbf{g}_1\mathbf{g}_2 \ldots \mathbf{g}_p$. The multiple indeterminate product is now taken and the space complement formed from every group of n corresponding factors. F_0 then appears as a sum of terms of the form

$$(\mathbf{a}_1\mathbf{b}_1 \ldots \mathbf{g}_1)\,(\mathbf{a}_2\mathbf{b}_2 \ldots \mathbf{g}_2) \ldots (\mathbf{a}_p\mathbf{b}_p \ldots \mathbf{g}_p). \tag{6}$$

Since the operation F_0 is distributively invariant the sum of all the terms of the form (6) and the sum of all the corresponding terms in the second complex are respectively the same functions of corresponding components. Also for each group of factors

$$(\mathbf{a}'\mathbf{b}' \ldots \mathbf{g}') = \epsilon(\mathbf{ab} \ldots \mathbf{g}) \tag{7}$$

and there are p groups in each term. Hence

$$\overline{F}_0((\mathbf{A}', \mathbf{B}', \ldots \mathbf{G}')) = \epsilon^p F_0((\mathbf{A}, \mathbf{B}, \ldots \mathbf{G})) \tag{8}$$

which proves the theorem and shows that the weight of the invariant is in this case equal to the degree of the polyadics.

Corollary to theorem I. Given n polyadics of degree p, the full sign scalar determinant of class $p+1$ whose layers of the first direction are the respective sets of n^p scalar components is an invariant of these polyadics.

This result, which enables us to write down immediately the invariants contemplated in theorem I, is an immediate consequence of R & H, p. 185. For it is there shown that the determinant in question is equal to F_0.

Example 1. Given two binary dyadics $\mathbf{e}_i\mathbf{e}_j a_{ij}$ and $\mathbf{e}_i\mathbf{e}_j b_{ij}$, the cubic determinant.

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} \begin{vmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{vmatrix} \tag{9}$$

which develops into

$$a_{11}b_{22}-a_{21}b_{12}-a_{12}b_{21}+a_{22}b_{11} \tag{10}$$

is an invariant of the two dyadics.

Example 2. Given two binary quadrics $x_ix_ja_{ij}$ and $x_ix_jb_{ij}$ with $a_{12}=a_{21}$ and $b_{12}=b_{21}$, the quantity (10) becomes

$$a_{11}b_{22}-2a_{12}b_{12}+a_{22}b_{11}, \tag{11}$$

an invariant of the two quadrics, as is well known.

This example illustrates the italicized remark under Article 4.

Example 3. Let $\mathbf{A}$ be a binary triadic $\mathbf{e}_i\mathbf{e}_j\mathbf{e}_k a_{ijk}$ and form another triadic $\mathbf{B}$ by permuting the factors of $\mathbf{A}$ so as to give $\mathbf{e}_k\mathbf{e}_i\mathbf{e}_j a_{ijk}$, one of the conjugates of $\mathbf{A}$. Fold $\mathbf{A}$ into $\mathbf{B}$ on all factors. The resulting invariant is the full sign determinant of four directions

$$\left|\begin{array}{cc|cc} a_{111} & a_{112} & a_{211} & a_{212} \\ a_{121} & a_{122} & a_{221} & a_{222} \\ \hline a_{111} & a_{121} & a_{112} & a_{122} \\ a_{211} & a_{221} & a_{212} & a_{222} \end{array}\right| \tag{12}$$

which develops into

$$\begin{aligned} &a_{111}a_{222}-a_{121}a_{122}-a_{112}a_{212}+a_{122}a_{112} \\ &-(a_{111}a_{222}-a_{211}a_{212}-a_{121}a_{221}+a_{221}a_{211}) \end{aligned} \tag{13}$$

that is into

$$a_{121}(a_{221}-a_{122})+a_{112}(a_{122}-a_{212})-a_{211}(a_{211}-a_{221}) \tag{14}$$

This invariant of the two triadics **A** and **B** is also an invariant of **A**. For it is a function of the components of **A**; and, in passing to the second complex, it does not matter whether we accent factors before or after permuting them, that is **B′** is conjugate to **A′** by the same permutation as **B** to **A**.

It may be noted that the invariant (14) of **A** changes sign when any pair of indices is interchanged throughout, and therefore vanishes if any pair be commutative.

Example 4. By the same reasoning as in Example 3 it may be shown that, given any number q of polyadics less than n, and forming $n-q$ polyadics each of which is either equal to or conjugate to one of the given polyadics, the result of the operation F_0 upon the resulting n polyadics is an invariant of the given q polyadics, — provided this result does not vanish identically.

Example 5. If p is even, the full sign determinant of class p whose elements are the scalar components of a polyadic of degree p is an invariant of the polyadic. For (by OD, p. 234, theorem IX) this determinant can be obtained by folding the polyadic into itself as $F_0((\mathbf{A}^n))$, aside from a constant scalar factor.

Example 6. If p is even, the full-sign determinant of class p whose elements are $a_{i_1 i_2 \cdots i_p}$ with all indices commutative is an invariant of the quantic $a_{i_1 i_2 \cdots i_p} x_{i_1} x_{i_2} \cdots x_{i_p}$. This is, like Example 2, an application of the italicized remark of Article 4. Thus every quantic of even degree (and every n) has an invariant of degree n. The fact that such invariants can be written as p-way determinants was shown by Zehfuss.

Example 7. If p is odd, a full sign determinant of class p whose elements are the scalar elements of a polyadic of degree p is *not* an invariant of the polyadic. It will be sufficient to prove this for a binary triadic, taking the first index in the resulting determinant as non-signant. The proof in all cases is similar. If the elements of the determinant are a_{ijk} those of the corresponding determinant in the second complex will be $\epsilon_{ui}\epsilon_{vj}\epsilon_{wk}a_{uvw}$. If the first direction be non-signant the coefficient of $a_{ijk}a_{rst}$ in the new determinant is found by actual expansion to be

$$(\epsilon_{i1}\epsilon_{r2}+\epsilon_{i2}\epsilon_{r1})(\epsilon_{j1}\epsilon_{s2}-\epsilon_{j2}\epsilon_{s1})(\epsilon_{k1}\epsilon_{t2}-\epsilon_{k2}\epsilon_{t1}) \qquad (15)$$

whence it is evident that the new determinant does not factor

into ϵ^s times the original determinant. This is due to the first factor of (15), which is a permanent instead of a determinant because the first direction is non-signant. If the sign of $\epsilon_{i2}\epsilon_{j1}$ were minus instead of plus clearly the conditions for an invariant would be satisfied.

6. Multiple invariants.

Let the process of converting a polyadic $\mathbf{A}$ into its corresponding polyadic $\mathbf{A}'$ (*i.e.* accenting its vector factors) be carried out in p successive steps, each step consisting of the replacement of one set of vector factors by their corresponding vectors in the second complex. We might replace all first factors, that is accent them, then all second factors, etc., up to all p^{th} or last factors, but evidently the result of any number of such replacements is independent of the order in which they are performed. Let m be an integer and suppose the m^{th} factors of $\mathbf{A}$ to be replaced. By (1), if $\mathbf{A}'_{(m)}$ denote the result of this replacement,

$$\mathbf{A}'_{(m)} = \mathbf{e}_{i_1}\mathbf{e}_{i_2}\cdots\mathbf{e}_{i_{m-1}}\mathbf{f}_{i_m}\mathbf{e}_{i_{m+1}}\cdots\mathbf{e}_{i_p}a_{i_1 i_2\cdots i_{m-1}j_m i_{m+1}\cdots i_p}\epsilon_{j_m i_m} \quad (16)$$

where the right side may be said to be "in the second complex on the m^{th} factor only." Similarly if the m_1^{th} factors and the m_2^{th} factors are replaced the result may be denoted by $\mathbf{A}'_{(m_1 m_2)}$ and so on, such polyadics being in the second complex on certain factors only. I shall also say that $\mathbf{A}'_{(m)}$ *corresponds to* $\mathbf{A}$ *on the* m^{th} *factor only*, and similarly for all cases.

Definition II. Let F be a scalar function of the scalar components of a set of given polyadics in the first complex, all of degree p. Let $\overline{\mathrm{F}}_{(m_1, m_2, \cdots m_q)}$ be the same function of the scalar components of those new polyadics which correspond to the given polyadics only on the q chosen factors denoted by the q integers $m_1, m_2, \ldots m_q$. If

$$\overline{\mathrm{F}}_{(m_1, m_2, \cdots m_q)} = \epsilon^n \mathrm{F} \quad (17)$$

the quantity F is said to be an invariant of the given polyadics on the $m_1, m_2, \ldots m_q$ *simultaneously*.

Theorem II. If F is an invariant of a set of polyadics on the m^{th} factor, and also invariant on the t^{th} factor, it is invariant on these two factors simultaneously.

Proof. By hypothesis

$$\overline{\mathrm{F}}_{(m)} = \epsilon^{\mu}\mathrm{F} \tag{18}$$

and also

$$\overline{\mathrm{F}}_{(t)} = \epsilon^{\tau}\mathrm{F} \tag{19}$$

where μ and τ are the weights of F on the respective factors. Now (18) and (19) are identities. We may replace, on both sides of (19), the scalar components of the given polyadics, each to each, by the scalar components of those polyadics which correspond to them on the m^{th} factor only, giving

$$\overline{\mathrm{F}}_{(m,\,t)} = \epsilon^{\tau}\overline{\mathrm{F}}_{(m)} \tag{20}$$

whence by (18)

$$\overline{\mathrm{F}}_{(m,\,t)} = \epsilon^{\tau}\epsilon^{\mu}\mathrm{F} \tag{21}$$

Corollary to theorem II. If F is invariant on any number of factors separately it is invariant on these factors simultaneously.

It is important to notice that the converse is not true.

Definition III. Given a set of polyadics of degree p, a scalar function of their scalar components which is invariant on each factor separately will be called a *multiple invariant*.

By the above corollary, a multiple invariant is an invariant. Since the converse of the corollary is not true, not all invariants are multiple invariants.

Example 1. The invariants of theorem I are multiple invariants.

For suppose only the first factors of the given polyadics to be replaced by their corresponding vectors. Let F_0 be written as a sum of terms like (6). Then $\overline{\mathrm{F}}_{0(1)}$ may be written as a sum of terms of the form

$$(\mathbf{a}_1'\mathbf{b}_1' \,.\,.\, \mathbf{g}_1')(\mathbf{a}_2\mathbf{b}_2 \,.\,.\, \mathbf{g}_2) \,.\,.\,.\, (\mathbf{a}_p\mathbf{b}_p \,.\,.\, \mathbf{g}_p) \tag{22}$$

whence

$$\overline{\mathrm{F}}_{0(1)} = \epsilon\mathrm{F}_0 \tag{23}$$

Example 2. If F is a polynomial in the components of the given polyadics, and is invariant on some one factor singly, the degree of F in all the components together cannot be less than n.

For by (16) the components of the polyadics which correspond to the given polyadics on one factor only are linear in the ϵ_{ji}. Hence if $\overline{F}_{(m)}$ contains ϵ as a factor it must be of degree n, at least. Under the same hypotheses it is evident that F must be homogeneous of degree μn where μ is an integer.

Example 3. The invariant $a_{12}-a_{21}$ of a binary dyadic is not a multiple invariant.

For if so it would be invariant on any factor singly by definition, which is impossible since it is of degree less than 2.

Example 4. The invariant (14) of the binary triadic **A** of Article 5, Example 3, is not a multiple invariant when considered as an invariant of **A** alone.

For, with the notation of that example, we may suppose **A** to be written as a sum of terms

$$\mathbf{A}=\mathbf{a}_1\mathbf{a}_2\mathbf{a}_3+\mathbf{b}_1\mathbf{b}_2\mathbf{b}_3+\ldots \tag{24}$$

while the conjugate **B** may be written

$$\mathbf{B}=\mathbf{a}_3\mathbf{a}_1\mathbf{a}_2+\mathbf{b}_3\mathbf{b}_1\mathbf{b}_2+\ldots \tag{25}$$

whence F_0 ((**A**, **B**)) may be written

$$(\mathbf{a}_1\mathbf{a}_3)\,(\mathbf{a}_2\mathbf{a}_1)\,(\mathbf{a}_3\mathbf{a}_2)+(\mathbf{a}_1\mathbf{b}_3)\,(\mathbf{a}_2\mathbf{b}_1)\,(\mathbf{a}_3\mathbf{b}_2)+(\mathbf{b}_1\mathbf{a}_3)\,(\mathbf{b}_2\mathbf{a}_1)\,(\mathbf{b}_3\mathbf{a}_2)+\text{etc.} \tag{26}$$

and $\overline{F}_{0(1)}$, which, by definition, is obtained by accenting only those vectors which are first factors of **A** (indicated by subscripts), may be written

$$(\mathbf{a}_1'\mathbf{a}_3)\,(\mathbf{a}_2\mathbf{a}_1')\,(\mathbf{a}_3\mathbf{a}_2)+(\mathbf{a}_1'\mathbf{b}_3)\,(\mathbf{a}_2\mathbf{b}_1')\,(\mathbf{a}_3\mathbf{b}_2)+\text{etc.} \tag{27}$$

It is now impossible that we have identically

$$\overline{F}_{0(1)}=\epsilon F_0 \tag{28}$$

For let the components of **A** be so chosen in numerical value that **A** reduces to the single triad $\mathbf{a}_1\mathbf{a}_2\mathbf{a}_3$. Then (27) reduces to its first term. Thus if (28) holds identically ϵ must be a factor of the first term of (27), manifestly not the case. Hence F_0((**A**, **B**)) is not a multiple invariant of **A** alone.

This invariant is, however, by Example 1 above, a multiple invariant of **A** and **B**, *i.e.* if we accent the first factors of **A** in (24) and the first factors of **B** in (25), and similarly for the other factors. This, however, implies regarding the components of **B** as different

from those of **A** . They are so if we compare them each to each instead of as a group.

Example 5. By similar reasoning it may be shown that the invariants of a set of q polyadics obtained in Article 5, Example 4, are invariant on those factors separately, and those only, which have remained unpermuted in the formation of conjugates.

7. **A method for finding invariants of a system of quantics.**

By the development (21′) of OD, page 224, a set of n^{p-k} k-adics appear as coefficients in the k-adic expansion of a polyadic of degree p. For brevity these k-adics may be spoken of as the k-adic coefficients on the last k factors.

Theorem III. An invariant F of a polyadic **P**, which is invariant on the last k factors simultaneously, is an invariant of the n^{p-k} k-adic coefficients on those factors.

Proof. By definition, when the last k factors *only* are accented, the k-adic coefficients are replaced by their corresponding k-adics in the second complex. Since F is invariant under these circumstances, by hypothesis, it is an invariant of these k-adics.

Corollary to theorem III. A multiple invariant of a polyadic is an invariant of any set of k-adic coefficients.

If we now let the last k indices be commutative, *i.e.*, suppose any scalar component of **P** to be unaltered in numerical value when the last k indices are permuted in any manner, we may pass from the k-adics to their related quantics of like degree. By definition, if F is an invariant of the k-adics it is an invariant of the quantics under the assumed commutativity of indices. (Or we may say that the k-adics are symmetrical, or that **P** is symmetrical on the last k indices.)

But it is important to notice that the converse of theorem III does not hold: an invariant of a set of k-adic coefficients in a k-adic expansion of **P** need not be an invariant of **P**. For example, with $n=2$, the cubic determinant whose elements are a_{ijk} with i non-signant is an invariant of the two binary dyadics $\mathbf{e}_j\mathbf{e}_k a_{ijk}$ and $\mathbf{e}_j\mathbf{e}_k a_{2jk}$, by the corollary to theorem I. But this determinant is not an invariant of the triadic $\mathbf{e}_i\mathbf{e}_j\mathbf{e}_k a_{ijk}$, by Article 5, Example 7, which shows that the determinant is invariant on the indices j and k but not on i.

8. **Second method for finding invariants of a system of quantics.**

Theorem IV. If a polyadic **A** is the direct product of two polyadics **B** and **C** and if F is an invariant of **A**, then F is also an invariant of **B** and **C**.

Proof. Since $\mathbf{A}=\mathbf{BC}$ the scalar components of **A** are functions of those of **B** and **C**. Hence F is a function of the scalar components of **B** and **C**. Passing to the second complex by accenting all vectors, we have $\mathbf{A}'=\mathbf{B}'\mathbf{C}'$, that is the components of $\mathbf{A}'$ are the same functions of those of $\mathbf{B}'$ and $\mathbf{C}'$ as if the accents were lacking. Therefore $\overline{\mathrm{F}}$ is the same function of the components of $\mathbf{B}'$ and $\mathbf{C}'$ that F is of those of **B** and **C**. Since $\overline{\mathrm{F}}=\epsilon^{n}\mathrm{F}$ by hypothesis, F is an invariant of **B** and **C**.

The same reasoning applies to an invariant of several polyadics some or all of which are expressible as indeterminate products of two or more factors. By finally replacing the polyadics by their related quantics, we have an invariant of these quantics.

Example 1. The simplest illustration is the invariant $a_{12}-a_{21}$ of a binary dyadic. Let the dyadic be $(\mathbf{e}_1a_1+\mathbf{e}_2a_2)\ (\mathbf{e}_1b_1+\mathbf{e}_2b_2)$ so that $a_{12}=a_1b_2$ and $a_{21}=a_2b_1$. The invariant becomes $a_1b_2-a_2b_1$, an invariant of the two vector factors, and of the linear quantics x_ia_i and x_ib_i.

Example 2. Given two binary dyadics $\mathbf{e}_i\mathbf{e}_ja_{ij}$ and $\mathbf{e}_i b_i\mathbf{e}_jb_j$. Folding gives the cubic determinant

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} \begin{vmatrix} b_1b_1 & b_1b_2 \\ b_2b_1 & b_2b_2 \end{vmatrix} \qquad (29)$$

which, if $a_{12}=a_{21}$, develops into

$$a_{11}b_2^2-2a_{12}b_1b_2+a_{22}b_1^2 \qquad (30)$$

an invariant (eliminant) of the quadric and linear quantics $x_ix_ja_{ij}$ and x_ib_i.

9. **Covariants of polyadics.**

Definition IV. Let **F** be a polyadic in the first complex whose scalar components are functions of the scalar components of a set of given polyadics in the first complex. Let $\overline{\mathbf{F}}$ be a polyadic in the second complex whose scalar components are the same functions of the scalar components of the corresponding polyadics. Let $\mathbf{F}'$ be the polyadic which corresponds to **F**. If $\overline{\mathbf{F}}=\epsilon^{n}\mathbf{F}'$ the polyadic **F** is said to be a covariant of the given polyadics.

As an example let the binary triadic $\mathbf{e}_i\mathbf{e}_j\mathbf{e}_k a_{ijk}$ be expressed as a sum of terms of the form $\mathbf{abc}$. Take space complements of the first two factors in every term, giving a sum of terms $(\mathbf{ab})\mathbf{c}$. The result is a covariant of the triadic. For if $\mathbf{F} = \Sigma(\mathbf{ab})\mathbf{c}$ we have $\mathbf{F}' = \Sigma(\mathbf{ab})\mathbf{c}'$ because $(\mathbf{ab})$ is a scalar; $\overline{\mathbf{F}} = \Sigma(\mathbf{a}'\mathbf{b}')\mathbf{c}'$ because the space complement is distributively invariant; and $\overline{\mathbf{F}} = \epsilon\mathbf{F}'$ because $(\mathbf{a}'\mathbf{b}') = \epsilon(\mathbf{ab})$ when $n = 2$. Applying this to the scalar expansion $\mathbf{e}_i\mathbf{e}_j\mathbf{e}_j a_{ijk}$ we obtain (cf. Article 4)

$$(a_{121} - a_{211})\mathbf{e}_1 + (a_{122} - a_{212})\mathbf{e}_2 + (a_{123} - a_{213})\mathbf{e}_3 \tag{31}$$

as a covariant of the triadic.

10. **Covariants obtained by folding.**

If polyadics are folded on all but the k last factors, the result is denoted by $\mathbf{F}_k$, or, if we wish to emphasize the fact that the result is a polyadic, by $\mathbf{F}_k$.

Theorem V. Given n polyadics $\mathbf{A}, \mathbf{B}, \ldots$ all of degree p, the polyadic $\mathbf{F}_k((\mathbf{A}, \mathbf{B}, \ldots))$ is a covariant of these polyadics,

Proof. With the notation of Article 5, $\mathbf{F}_k$ appears as a sum of terms of the form

$$(\mathbf{a}_1\mathbf{b}_1 \ldots \mathbf{g}_1)(\mathbf{a}_2\mathbf{b}_2 \ldots \mathbf{g}_2) \ldots (\mathbf{a}_{p-k}\mathbf{b}_{p-k} \ldots \mathbf{g}_{p-k})\, \mathbf{a}_{p-k+1}\mathbf{b}_{p-k+1} \ldots \mathbf{g}_{p-k+1} \ldots \mathbf{a}_p\mathbf{b}_p \ldots \mathbf{g}_p$$

and the remainder of the proof follows as in the last article, giving $\overline{\mathbf{F}}_k = \epsilon^{p-k}\mathbf{F}_k'$.

As an example let the binary triadic $\mathbf{e}_i\mathbf{e}_j\mathbf{a}_{ij}$ be folded into itself on the first two factors. The result (OD, Article 10, p. 225) is the determinant with vector elements and with non-signant layers alike

$$\begin{vmatrix} \mathbf{a}_{11} & \mathbf{a}_{12} \\ \mathbf{a}_{21} & \mathbf{a}_{22} \end{vmatrix} \begin{vmatrix} \mathbf{a}_{11} & \mathbf{a}_{12} \\ \mathbf{a}_{21} & \mathbf{a}_{22} \end{vmatrix} \tag{32}$$

which is a dyadic (OD, p. 235). If we replace $\mathbf{e}_i$ by x_i and make indices commutative the result is twice the quadric

$$x_1^2(a_{111}a_{122} - a_{112}^2) + x_1x_2(a_{111}a_{222} - a_{112}a_{122}) + x_2^2(a_{112}a_{222} - a_{221}^2) \tag{33}$$

a covariant (Hessian) of the binary cubic $x_ix_jx_ka_{ijk}$. On the other hand, if we allow only the *last two* indices to become commutative

when we replace $\mathbf{e}_i$ with x_i, (32) develops into twice the quadric

$$x_1^2(a_{111}a_{212}-a_{112}a_{211})+x_1x_2(a_{111}a_{222}-a_{122}a_{211})+x_2^2(a_{112}a_{222}-a_{122}a_{212}) \quad (34)$$

a covariant (Jacobian) of the quadrics $x_ix_ja_{1ij}$ and $x_ix_ja_{2ij}$. It may be noticed that we have here employed the principle of theorem III without stopping to restate it with reference to covariants.

11. **Contravariants of polyadics.**

Let $\bar{\epsilon}_{ij}$ be the cofactor of ϵ_{ij} in ϵ and let $\frac{\bar{\epsilon}_{ij}}{\epsilon}=E_{ij}$. To any chosen vector $\mathbf{a}$ in the first complex let a vector $\mathbf{a}''$ of the new complex be made to correspond by virtue of the vector equation

$$\mathbf{a}''=\mathbf{f}_ia_i''=\mathbf{f}_iE_{ji}a_j \quad (35)$$

I shall *not* say, however, that $\mathbf{a}''$ *corresponds* to $\mathbf{a}$, because it has already been defined that $\mathbf{a}'$ corresponds to $\mathbf{a}$. It appears convenient to say that $\mathbf{a}''$ is *complementary* to $\mathbf{a}$ in the second complex. A polyadic $\mathbf{A}''$ formed by replacing all vector factors of a given polyadic $\mathbf{A}$ by their complementary vectors in the second complex will be called *complementary to the given polyadic in the second complex*. The name "complementary" is suggested by the fact that, as will soon be proved, if two vectors, one in each complex, are related by (1), their space complements are related by (35). And more generally, given a polyad $\mathbf{ab}\ldots\mathbf{g}$ of degree less than n, the space complement $(\mathbf{a'b'}\ldots\mathbf{g'})$ in the second complex is a polyadic complementary to the polyadic $(\mathbf{ab}\ldots\mathbf{g})$ in the sense just defined, aside from the factor ϵ. This important relation may be expressed by the equation

$$(\mathbf{a'b'}\ldots\mathbf{g'})=\epsilon(\mathbf{ab}\ldots\mathbf{g'})' \quad (36)$$

Thus the space complement of the polyad is not a covariant, for if it were we should have a single instead of a double accent on the right side.

Definition V. Let $\mathbf{F}$ be a polyadic in the first complex whose scalar components are functions of the scalar components of a set of given polyadics in the first complex. Let $\overline{\mathbf{F}}$ be a polyadic in the second complex whose scalar components are the same functions of the scalar components of the corresponding polyadics. Let $\mathbf{F}''$

be the polyadic complementary to $\mathbf{F}$ in the second complex. If $\overline{\mathbf{F}} = \epsilon^n \mathbf{F}''$ the polyadic $\mathbf{F}$ is said to be a contravariant of the given polyadics.

12. **The space complement as the source of contravariants.**

I shall now prove (36). Let there be q vectors $\mathbf{a}, \mathbf{b}, \ldots, \mathbf{g}$ with $q < n$. By definition

$$(\mathbf{ab}\ldots\mathbf{g}) = \begin{vmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \ldots & \mathbf{e}_n \\ \mathbf{e}_1 & \mathbf{e}_2 & \ldots & \mathbf{e}_n \\ \ldots & \ldots & \ldots & \ldots \\ \mathbf{e}_1 & \mathbf{e}_2 & \ldots & \mathbf{e}_n \\ a_1 & a_2 & \ldots & a_n \\ b_1 & b_2 & \ldots & b_n \\ \ldots & \ldots & \ldots & \ldots \\ g_1 & g_2 & \ldots & g_n \end{vmatrix} \tag{37}$$

where the first $n-q$ rows are alike. We now form $(\mathbf{a'b'}\ldots\mathbf{g'})$ by replacing the $\mathbf{e}_i$ by the basis vectors of the second complex, namely the $\mathbf{f}_i$ and at the same time replacing the scalar components of the given vectors by those of their corresponding vectors as in (2). Thus

$$(\mathbf{a'b'}\ldots\mathbf{g'}) = \begin{vmatrix} \mathbf{f}_1 & \mathbf{f}_2 & \ldots & \mathbf{f}_n \\ \ldots & \ldots & \ldots & \ldots \\ \mathbf{f}_1 & \mathbf{f}_2 & \ldots & \mathbf{f}_n \\ \epsilon_{j1}a_j & \epsilon_{j2}a_j & \ldots & \epsilon_{jn}a_j \\ \epsilon_{j1}b_j & \epsilon_{j2}b_j & \ldots & \epsilon_{jn}b_j \\ \ldots & \ldots & \ldots & \ldots \\ \epsilon_{j1}g_j & \epsilon_{j2}g_j & \ldots & \epsilon_{jn}g_j \end{vmatrix} \tag{38}$$

In the third place we substitute $\mathbf{e}_i''$ for the $\mathbf{e}_i$ in (37), noting that for basis vectors (35) reduces to

$$\mathbf{e}_i'' = \mathbf{f}_k E_{ik} \tag{39}$$

whence we obtain

$$(\mathbf{ab}\ldots\mathbf{g})'' = \begin{vmatrix} \mathbf{f}_k E_{1k} & \mathbf{f}_k E_{2k} & \ldots & \mathbf{f}_k E_{nk} \\ \ldots & \ldots & \ldots & \ldots \\ \mathbf{f}_k E_{1k} & \mathbf{f}_k E_{2k} & \ldots & \mathbf{f}_k E_{nk} \\ a_1 & a_2 & \ldots & a_n \\ \ldots & \ldots & \ldots & \ldots \\ g_1 & g_2 & \ldots & g_n \end{vmatrix} \tag{40}$$

whence (36) follows by virtue of the identity

$$\epsilon_{ji}\mathbf{f}_k E_{jk}=\mathbf{f}_i \tag{41}$$

In the first part of this paper it was shown that invariants were formed from a polyadic, or from an indeterminate product of polyadics, by arranging the vector factors in groups, each group containing n vectors, and taking the space complement of each group, in that case a scalar. In a similar manner, if we arrange the vector factors in groups, some it may be containing n factors, but others containing less than n factors, or even a single vector, and take the space complement of all groups, (being careful to treat all terms in the same way so as to preserve distributive invariance) the result, because of the distributive invariance of the space complement (proved by Naess), must be a contravariant of the given polyadic or polyadics.

If some vector factors are left free or " unfolded " these factors will have covariant character, as already exemplified. The general case may be said to give a mixed concomitant of the given polyadics, by analogy with the usual terminology. For example if, in the proof of theorem V, we assume only q polyadics ($q<n$) we find that $\mathbf{F}_k((\mathbf{A}, \mathbf{B}, \ldots))$ is in general a mixed concomitant, which reduces to a contravariant in the case $\mathbf{F}_0$.

Example 1. Given a binary dyadic $\mathbf{e}_1\mathbf{e}_1a_{11}+\mathbf{e}_1\mathbf{e}_2a_{12}+\mathbf{e}_2\mathbf{e}_1a_{21}+\mathbf{e}_2\mathbf{e}_2a_{22}$, replace every vector factor by its space complement, noting that $(\mathbf{e}_1)=-\mathbf{e}_2=-\mathbf{E}_2$ and $(\mathbf{e}_2)=+\mathbf{e}_1=+\mathbf{E}_1$, where capitals are introduced merely to remind us of the contravariant character. The result is

$$\mathbf{E}_2\mathbf{E}_2a_{11}-\mathbf{E}_2\mathbf{E}_1a_{12}-\mathbf{E}_1\mathbf{E}_2a_{21}+\mathbf{E}_1\mathbf{E}_2a_{22} \tag{42}$$

which, as above shown, must be a contravariant of the dyadic. If we further make indices commutative and put $\mathbf{E}_1=\xi$ and $\mathbf{E}_2=\eta$ we have

$$\eta^2a_{11}-2\xi\eta a_{12}+\xi^2a_{22} \tag{43}$$

a contravariant of $x_ix_ja_{ij}$ in the usual sense.

13. **More general examples of folding.**

The folding process may now be employed in a somewhat freer manner by utilizing the familiar principle that an invariant of covariants of given forms is itself an invariant of the given forms.

Begin with four binary triadics expanded on their last factors as

$$\mathbf{A}=\mathbf{e}_r\mathbf{e}_s\mathbf{a}_{ij},\ \mathbf{B}=\mathbf{e}_r\mathbf{e}_s\mathbf{b}_{ij},\ \mathbf{C}=\mathbf{e}_r\mathbf{e}_s\mathbf{c}_{ij},\ \text{and}\ \mathbf{D}=\mathbf{e}_r\mathbf{e}_s\mathbf{d}_{ij} \tag{44}$$

Make the two dyadics

$$\mathbf{M}=\mathrm{F}_1((\mathbf{A},\mathbf{B}))\ \text{and}\ \mathbf{N}=\mathrm{F}_1((\mathbf{C},\mathbf{D})); \tag{45}$$

finally make the scalar

$$I=\mathrm{F}_0((\mathbf{M},\mathbf{N})) \tag{46}$$

If we think of the four triadics as sums of terms of the form

$$\mathbf{a}_1\mathbf{a}_2\mathbf{a}_3,\ \mathbf{b}_1\mathbf{b}_2\mathbf{b}_3,\ \mathbf{c}_1\mathbf{c}_2\mathbf{c}_3,\ \text{and}\ \mathbf{d}_1\mathbf{d}_2\mathbf{d}_3, \tag{47}$$

respectively, we may write

$$\mathbf{M}=(\mathbf{a}_1\mathbf{b}_1)\ (\mathbf{a}_2\mathbf{b}_2)\mathbf{a}_3\mathbf{b}_3\ \text{and}\ \mathbf{N}=(\mathbf{c}_1\mathbf{d}_1)\ (\mathbf{c}_2\mathbf{d}_2)\mathbf{c}_3\mathbf{d}_3 \tag{48}$$

omitting summation signs. Then I becomes

$$(\mathbf{a}_1\mathbf{b}_1)\ (\mathbf{a}_3\mathbf{b}_2)\ (\mathbf{c}_1\mathbf{d}_1)\ (\mathbf{c}_2\mathbf{d}_3)\ (\mathbf{a}_3\mathbf{c}_3)\ (\mathbf{b}_3\mathbf{d}_3) \tag{49}$$

and is therefore an invariant of the triadics. I shall now utilize (45) and (46) to express I in determinant form. We know that (OD, Article 10)

$$\mathbf{M}=\begin{vmatrix} \mathbf{a}_{11} & \mathbf{a}_{12} \\ \mathbf{a}_{21} & \mathbf{a}_{22} \end{vmatrix}\begin{vmatrix} \mathbf{b}_{11} & \mathbf{b}_{12} \\ \mathbf{b}_{21} & \mathbf{b}_{22} \end{vmatrix} \tag{50}$$

and similarly for $\mathbf{N}$. Let $\mathbf{M}=\mathbf{e}_r\mathbf{e}_s m_{rs}$. Then

$$m_{rs}=\begin{vmatrix} a_{11r} & a_{12r} \\ a_{21r} & a_{22r} \end{vmatrix}\begin{vmatrix} b_{11s} & b_{12s} \\ b_{21s} & b_{22s} \end{vmatrix} \tag{51}$$

(a case of OD (39) p. 231) and similarly for n_{rs} by putting c for a and d for b. We also know that (R & H)

$$I=\begin{vmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{vmatrix}\begin{vmatrix} n_{11} & n_{12} \\ n_{21} & n_{22} \end{vmatrix} \tag{52}$$

We thus have I as a cubic determinant whose elements are themselves cubic determinants. Since a cubic determinant expands into four terms, there are 64 terms in the development, which is linear in the four sets of eight components each.

If all four triadics are equal, $m=n$ and $I_a=2(m_{11}m_{22}-m_{12}m_{21})$ but also $a=b$, hence $m_{12}=m_{21}=a_{111}a_{222}-a_{211}a_{122}-a_{122}a_{212}+a_{221}a_{112}$

while $m_{11}=2\,(a_{111}a_{221}-a_{121}a_{211})$ and $m_{22}=2\,(a_{112}a_{222}-a_{122}a_{212})$. Aside from the factor -2 we thus have

$$\begin{aligned} I_a &= (a_{111}a_{222}-a_{211}a_{122}-a_{121}a_{212}+a_{221}a_{112})^2 \\ &\quad -4(a_{111}a_{221}-a_{121}a_{211})\,(a_{112}a_{222}-a_{122}a_{212}) \\ &= a^2_{111}a^2_{222}+a^2_{112}a^2_{221}+a^2_{212}a^2_{121}+a^2_{211}a^2_{122} \\ &\quad -2(a_{111}a_{222}a_{112}a_{221}+a_{111}a_{222}a_{212}a_{121}+a_{111}a_{222}a_{211}a_{122} \\ &\quad\ +a_{112}a_{221}a_{212}a_{121}+a_{112}a_{221}a_{211}a_{122}+a_{212}a_{121}a_{211}a_{122}) \\ &\quad\ +4(a_{111}a_{122}a_{212}a_{221}+a_{112}a_{121}a_{211}a_{222}) \end{aligned} \tag{53}$$

which contains 24 terms and is symmetrical in the three indices, in the sense that, if any pair of indices be interchanged throughout, the expression is unaltered as a whole, though individual terms may be altered.

In view of the symmetrical character of I_a in the three factors of $\mathbf{A}$, inspection of (49) shows that we might have obtained the same result by a different style of folding. Make the tetradic $F_2((\mathbf{A}, \mathbf{A}))$, *i.e.*, fold on the first factor only, and call this tetradic $\mathbf{P}$. The scalar components of $\mathbf{P}$ are two-way determinants or else zero (OD, Article 11). Then $I_a=F_0((\mathbf{P}, \mathbf{P}))$, as a four-way determinant whose elements are two-way determinants or zero. It is easy to verify by actual expansion that the result agrees with (53).

Again, further inspection of (49) shows that we may obtain I_a as a seven-way determinant whose elements are products of two scalar components as follows: take the indeterminate (direct) product of $\mathbf{A}$ into itself, a hexadic $\mathbf{H}=\mathbf{AA}=\mathbf{a}_1\mathbf{a}_2\mathbf{a}_3\mathbf{b}_1\mathbf{b}_2\mathbf{b}_3$. Make a new hexadic by interchanging the third and sixth factors, $\mathbf{H}_1=\mathbf{c}_1\mathbf{c}_2\mathbf{d}_3\mathbf{d}_1\mathbf{d}_2\mathbf{c}_3$. Then make $F_0((\mathbf{H}, \mathbf{H}_1))$.

The triadic $\mathbf{A}$ can be expressed as the sum of *two* triads when I_a is *not* zero. For let $\mathbf{A}=\mathbf{a}_1\mathbf{a}_2\mathbf{a}_3+\mathbf{b}_1\mathbf{b}_2\mathbf{b}_3$. Then $\mathbf{M}=\mathrm{c}[\mathbf{a}_3\mathbf{b}_3+\mathbf{b}_3\mathbf{a}_3]$ where c is a scalar. But I_a is the determinant of the dyadic $\mathbf{M}$. Unless this invariant is zero $\mathbf{M}$ can be reduced to the form just written by an easy calculation, hence $\mathbf{A}=\phi\mathbf{a}_3+\theta\mathbf{b}_3$, where ϕ and θ are dyads.

Again, (49) shows that I_a is a *multiple* invariant. Hence it is an invariant of the two dyadics $\mathbf{e}_i\mathbf{e}_j a_{1ij}$, $\mathbf{e}_i\mathbf{e}_j a_{2ij}$, by theorem III. It is easy to verify that it is the eliminant of their related quadrics if the last two indices be made commutative. Finally if all indices are commutative it reduces to the discriminant of a binary cubic.

14. Conclusion.

The examples given in this paper, dealing with binary forms, have been chosen to illustrate as briefly as practicable the three peculiarities mentioned in Article 1. There is clearly nothing whatever in the nature of the processes employed to impose this restriction. In the sequel it is planned to present certain features of the method in a more general way, and to show more fully its relation to the usual symbolic method. What has been so far accomplished, in the logic of the subject, consists mainly in the fact that an invariant of a polyadic system, such as (52), produces a variety of invariants of quantics when various special conditions are imposed on the given polyadics. As a working method, it is clear that the folding process, yielding determinants of higher class, enables us to write down such invariants or other concomitants in numerous ways.

In part II it will be shown that any concomitant rational and integral in the scalar components of a set of given polyadics, or in the coefficients of a set of given forms, is a special case of a concomitant *linear* in the components of a polyadic of higher degree. In other words the method of Art. 8 Ex. 1, is universal. From this fact the relation of the present method to the usual symbolic method follows readily.